An Investigation into Novel Helicases from *Plasmodium falciparum*

Luke Shaun Evans

The thesis is submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of Portsmouth

September 2012

ABSTRACT

With 800 thousand deaths each year resulting from more than 225 million cases of malaria worldwide, the increase of drug-resistant parasites and insecticide-resistant *Anopheles gambiae* mosquitoes has given rise to the call for new approaches in combating malaria. One proposed strategy focuses on the targeting of *Plasmodium* helicases due to their necessity in all cellular processes, and therefore the survival of the parasite and the infection of their host.

This study describes the novel purification and storage strategies of two *Plasmodium falciparum* helicases; PFEIF4A (formally PFH45) and PFDH60. This resulted in the yield of relatively large amounts of protein that can be used in high-throughput screening of inhibitors and structural-guided drug design in the future.

Moreover, further biochemical and biophysical characterisation has been performed to investigate the specific substrate requirements of each helicase, and the optimal conditions for helicase activity. As well as elucidating information that could provide pathways for specific targeting by both established and novel therapeutic molecules, particularly in the case of PFEIF4A, exposure to temperature just 3 °C above its 37 °C optimum has been shown to have an irreversible and detrimental effect to the proteins activity *in vitro*.

Preliminary experiments into the proteins' structures have been performed, as well as experiments exploring the effects of a novel DNA-interactive PBD conjugate and the generation and characterisation of aptamers to PFEIF4A. These experiments have contributed to future investigations in to novel anti-malarial therapeutics and diagnostic platforms.

CONTENTS LIST

DECLARATION

Whilst registered on the above degree I have not been registered for any other research award. The results and conclusions that are embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Word count = 37 738

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

A

4ABA	PFEIF4A Binding Aptamer
A	Adenine
AAA$^+$	ATPases Associated with various cellular Activities
AAV	Adeno-Associated Virus
ACTs	Artemisinin-based Combination Treatments
AMP-PnP	Adenosine 5'-(β,γ-imido)triphosphate
APS	Ammonium Persulphate
ASP	Active Site Pressurisation
ATPγS	adenosine 5'-O-(3-thio)triphosphate
ATRX	Alpha Thalassemia/mental Retardation syndrome X-linked
AUC	Analytical Ultracentrifuge

B

BSA	Bovine Serum Albumin

C

C	Cytosine
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
Cu	Curie(s)

D

DAPI	4’,6-Diamidino-2-Phenylindole
Dda	DNA-dependant ATPase
DDX; 4, 5	DEAD Box polypeptide; 4, 5
DEAE	Diethylaminoethyl
DHX16	DEAH-box peptide 16
DLS	Dynamic Light Scattering
DMF	Dimethylformaldehyde
DNA	Deoxyribonucleic Acid
DnaB	replicative DNA helicase B
dNTP	Deoxynucleotide Triphosphate
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
DTT	Dithiothreitol

E

EIF; 4A, 4E, 4G	Eukaryotic Initiation Factor; 4A, 4E, 4G
EMMPRIN	Extracellular Matrix Metalloproteinase Inducer
EMSA	Electophoretic Mobility Shift Assay

G

G	Guanine
gp4	Gene 4 protein
GST	Glutathione S-Transferase

H

HBA	HsdR Binding Aptamer
HCV	Hepatitis C Virus
HEF	Helicase-associated Endonuclease for Fork-structured DNA
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulphonic acid
His_6	Hexa-Histidine
HIV	Human Immunodeficiency Virus
HPD	Hairpin Duplex
HPV18	Human PapillomaVirus serotype 18
HRP	Horseradish Peroxidase
HSQC	Heteronuclear Single Quantum Coherence
HSV-1	Herpes Simplex Virus type 1

I

IgG	Immunoglobulin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IPTi	Intermittent Preventative Treatment for infants
IPTp	Intermittent Preventative Treatment in pregnancy

IRS	Indoor Residual Spraying
ITNs	Insecticidal Treated Nets

K

K_d	Dissociation Constant

L

LB	Lysogeny Broth

M

MBA	Mtase Binding Aptamer
MCM	Mini Chromosome Maintenance
MES	2-(N-Morpholino)ethylenesulphonic acid
Mfd	Mutation frequency decrease
M_r	Molecular Weight
Mtase	Methylase
MWCO	Molecular Weight Cut-Off

N

NADH	Reduced β-Nicotinamide adenine dinucleotide
NDP	Nucleotide Diphosphate
NER	Nuclear Excision Repair

NF-Y	Nuclear Factor Y
NMR	Nuclear Magnetic Resonance
NS3	Non-Structural protein 3
NTP	Nucleotide Triphosphate

P

PAGE	Polyacrylamide Gel Electrophoresis
PBD	Pyrrolobenzodiazepine
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PcrA	Plasmid copy reduced A
PDB	Protein Database
PEP	Phosphoenolpyruvate
PFDH60	*P. falciparum* DEAD-box Helicase 60 kDa
PFEIF4A	*P. falciparum* EIF4A
PFEMP1	*P. falciparum* Erythrocyte Membrane Protein 1
PFH45	*P. falciparum* Helicase 45 kDa
PFPABP	*P. falciparum* Poly-A Binding Protein
PFRH5	*P. falciparum* Reticulocyte binding protein Homologue 5
PFSSB	*P. falciparum* Single-Stranded Binding protein
PHYRE	Protein Homology/Analogy Recognition Engine
P_i	Inorganic Phophate

Pif1	Phytochrome interacting factor 1
PK	Pyruvate kinase
PKC	Protein Kinase C
PMSF	Phenylmethanesulphonylfluoride
PP	PreScission Protease
ppm	Parts-per-million
pRBC	parasitised RBC
PVA	Polyvinyl Alcohol

R

R_0	Rotational radius
RBC	Red Blood Cell
Rec; A, B, C, D, Q	Recombination protein; A, B, C, D, Q
Rep	Rab escort protein
R_g	Radius of gyration
R_h	hydrodynamic radius
Rho	Rhodopsin
RM	Restriction Modification
RNA	Ribonucleic Acid
Rrm3	RNA recognition motif 3
rRNA	ribosomal RNA
RT	Reverse Transcriptase

Ruv; A, B, C	Resistance to UV; A, B, C

S

SAXS	Small Angle X-ray Scattering
SDS	Sodium Dodecyl Sulphate
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SF; 1, 2, 3, 4, 5, 6	Superfamily; 1, 2, 3, 4, 5, 6
Snf2	Sucrose non-fermenting 2
SP	Sulphadoxine-Pyrimethamine
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
SV40	Simian Virus 40

T

T	Thrombin
T	Thymine
TEMED	Tetramethylethylenediamine
TFIIH	Human Transcription Factor II complex
TFO	Triplex Forming Oligonucleotide
TRYR	Trypanothione Reductase

U

U	Uracil
UV	Ultraviolet
UvrD	UV resistant D

W

WHO	World Health Organisation
WRN	Werner

X

XPD	Xeroderma Pigmentosum D

Units

°	Degree(s)
A	Ampere(s)
Å	Angstrom(s)
aa	Amino acid(s)
AU	Absorbance Unit(s)
bp	Base pair(s)
C	Celsius/Centigrade
Da	Dalton(s)
eV	Electronvolt(s)
g	Gram(s)

K	Kelvin
L	Litre(s)
m	Metre(s)
M	Molar
mol	Mole(s)
nt	Nucleotide(s)
Pa	Pascal(s)
S	Svedberg(s)
V	Volt(s)

Metric Prefixes

k	Kilo-	$\times 10^{3}$
c	Centi-	$\times 10^{-2}$
m	Milli-	$\times 10^{-3}$
µ	Micro-	$\times 10^{-6}$
n	Nano-	$\times 10^{-9}$
p	Pico-	$\times 10^{-12}$
f	Femto-	$\times 10^{-15}$

ACKNOWLEDGMENTS

I would firstly like to thank my supervisors past and present Dr Keith Firman and Dr Darren Gowers, for giving me the opportunity to further my passion of research. I would like to make special mention to Dr James Youell, whose enthusiasm and forthcoming attitude has inspired me and helped me to stay focused.

I would also like to take the opportunity to thank everyone who has helped to make the process of carrying out my studies as enjoyable and fulfilling as they have been. Particular mention to my contemporaries Louise, Jack, and Charlotte, my senior colleagues Andy and Anastasia, the technical staff with special mention to Richard, and my collaborative colleagues most notably Irfaan.

I would like to thank my family and friends for their love and support throughout this time. Finally, I would like to dedicate this thesis to my fiancée Nicola for her enduring love, support, and patience, which has fuelled my motivation and ambition throughout the whole process.

DISSEMINATION OF RESEARCH

Publications

"A Synthetic Biology Project - Developing a single-molecule device for screening drug-target interactions." FEBS Lett **586**(15): 2157-2163.

Keith Firman; **Luke Evans**; James Youell

This review describes a European-funded project in the area of Synthetic Biology. The project seeks to demonstrate the application of engineering techniques and methodologies to the design and construction of a biosensor for detecting drug–target interactions at the single-molecule level. Production of the proteins required for the system followed the principle of previously described "bioparts" concepts (a system where a database of biological parts – promoters, genes, terminators, linking tags and cleavage sequences – is used to construct novel gene assemblies) and cassette-type assembly of gene expression systems (the concept of linking different "bioparts" to produce functional "cassettes"), but problems were quickly identified with these approaches. DNA substrates for the device were also constructed using a cassette-system. Finally, micro-engineering was used to build a magnetoresistive Magnetic Tweezer device for detection of single molecule DNA modifying enzymes (motors), while the possibility of constructing a Hall Effect version of this device was explored. The device is currently being used to study helicases from Plasmodium as potential targets for anti-malarial drugs, but we also suggest other potential uses for the device.

"Enhanced purification and characterization of the PfeIF4A (PfH45) helicase from Plasmodium falciparum using a codon-optimised clone." Protein Expr Purif **85**(1): 1-8.

Luke Evans; Darren Gowers; Keith Firman; James Youell

With the intention of investigating the DNA strand displacement properties of Plasmodium falciparum helicase PfeIF4A (formerly known as PfH45) a codon-optimized gene for

expression in Eschrrichia coli has been produced. Several histidine-containing proteins with intrinsic helicase activity were captured from the bacterial sonicate by initial Ni^{2+}-chromatography. Heparin and size-exclusion steps were subsequently required for unambiguous PfeIF4A purification. This strategy generated an active recombinant protein of significantly improved yield in comparison to previously published studies (approximately 4.2 mg/g wet weight of cells). Helicase unwinding assays confirmed a bipolar activity, but revealed a preference for unwinding a free 3'-end, with a rate of displacement in the 3'-5' direction 2-fold higher than that in the 5'-3' direction. DNA constructs with two, three or four blunt ends were not unwound. Studies confirmed the enzyme to be Mg^{2+}-dependent, optimally active at 37 °C and had a background ATP turnover rate of 23.16 ± 1.74 pmol/min, which in the presence of single- or double-stranded DNA doubled to 42.92 ± 3.21 pmol/min.

Oral Presentations

Characterisation of Novel Helicases from *Plasmodium falciparum*

Presented by **Luke Evans** at the Institute of Biomolecular and Biomedical Sciences (IBBS) conference at the University of Portsmouth (2012)

Poster Presentations

The Biochemical and Structural Characterisation of Novel Helicases from *Plasmodium falciparum*

Presented by **Luke Evans** at the Federation of American Societies for Experimental Biology (FASEB) Helicase conference at Steamboat Springs, Colarado, USA (2011)

CHAPTER ONE – INTRODUCTION

1.1 Helicases

Translocases are nucleic acid motor proteins that hydrolyse adenosine triphosphate (ATP) upon binding to deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) molecules. They transform energy derived from ATP hydrolysis to move directionally along the phosphodiester backbone. Helicases are a subclass of translocases, capable of unwinding the nucleic acid duplex (DNA, RNA, DNA-RNA hybrids) into its component strands as they move along it, a function which separates them from other translocases. As a consequence they are vital to all living organisms, playing a role in key cellular processes, such as DNA repair and replication (Cobb *et al.* 2002), transcription (Ciampi 2006), translation (Rogers *et al.* 1999), recombination (Singleton *et al.* 2004), RNA maturation and splicing (Staley *et al.* 1998), RNA degradation (Py *et al.* 1996), ribosome biogenesis (Venema *et al.* 1995), nuclear export processes (Kanai *et al.* 2004), and mitochondrial genome replication (Korhonen *et al.* 2003).

1.2 Biochemical Activities of Helicases

Helicase activity is a complex process with several biochemical properties that influence the many aspects that make up its mechanism. Singleton *et al.* (2007) describe these activities in terms of rate, processivity, directionality, step size, and active or passive mechanisms. Rate is defined as the number of base pairs unwound or the number of ATP molecules hydrolysed per unit of time. Processivity is the number of base pairs unwound per substrate binding event. Directionality is used to describe a processive helicases' bias in its movement along a nucleic acid strand (3' to 5' or 5' to 3'). Step size can describe the average distance moved (mechanical step size) or the average number of base pairs unwound (kinetic step size) per catalytic cycle. Finally, in addition to the conventional sense of an active or passive enzyme; differentiating whether or not an enzyme derives energy from the hydrolysis of ATP. Active and passive is used to describe the initial mechanism for helicase activity.

1.2.1 Rate

Many studies have focused on helicase rate of translocation, resulting in observations from a few to several thousand base pairs per second (bp s^{-1}) under comparable conditions. For example, monomers of the T7 bacteriophage 2 DNA primase (Kim *et al.* 2002), and the Herpes Simplex Virus type 1 (HSV-1) primase (Falkenberg *et al.* 1998) have shown rates of approximately 130 and 60 bp s^{-1} respectively.

Alternative studies have shown that rate can be affected and is likely controlled by many contributory factors. For instance, the translocation rate of 130 bp s^{-1} by T7 primase is based on the monomer interacting with single-stranded DNA (ssDNA). Whereas, translocation on double-stranded DNA (dsDNA) is only 10 bp s^{-1} unless coupled with a preceding DNA replisome activity, which catalyses the strand displacement (Stano *et al.* 2005). Eukaryotic Initiation Factor (EIF) 4A activity is also affected by other proteins. Alone eIF4A can only dissociate RNA duplexes of sufficiently low thermodynamic stability (Rogers *et al.* 1999). However, activity is observed to be up-regulated when in the presence of eIF4B as part of the eIF4 complex (Rogers *et al.* 2001; Marsden *et al.* 2006).

It makes sense that helicase activity would be affected by other factors as unregulated activity is likely to be devastating to the cell, which invests substantial effort in maintaining more stable dsDNA over less stable ssDNA.

1.2.2 Directionality

Due to the anti-parallel nature of duplex DNA it is important that motor proteins have a sense of direction. For proteins with an affinity for single strand substrates this is not a problem; once the protein is bound the polarity of DNA allows for directionality of translocation. dsDNA however is bipolar; two polar ssDNA molecules with the phosphodiester backbones anti-parallel to one another gives the molecule an intrinsic symmetry. Proteins that translocate along dsDNA appear to determine direction by tracking one of the two strands even if the protein is tracking the duplex.

Studies of the hexameric helicase of T7 primase illustrate this with a proposed binding change mechanism (Singleton *et al.* 2000) where the sequential association and dissociation of nucleotide triphosphates (NTP), nucleotide diphosphates (NDP), and inorganic phosphate (P_i) result in a cyclical transfer of the nucleic acid substrate between the monomers within the central cavity.

1.2.3 Processivity

Processivity is a measure of the average number of associated functions or catalytic cycles performed per substrate-binding event (number of base pairs unwound per molecule of ATP hydrolysed). As the translocases characteristic function is to move along nucleic acids, processivity is a major aspect of the mechanism. Therefore, to some degree the same is probably true for helicases as they are required to separate large duplex DNA molecules. As with rate of translocation it is likely that helicase processivity is greatly dependent on their interaction with other proteins.

1.2.4 Step Size

Aside from the average number of catalytic cycles per substrate-binding event (processivity), the number of base pairs unwound per catalytic cycle is also of interest. This is referred to as the step size and commonly accompanies each ATP hydrolysis event. Determining the step size is difficult due to factors such as base composition, supercoiling, and unusual structures such as cruciform and tetraplexes, however a theoretical step size can be calculated (Lohman *et al.* 1996).

1.2.5 Active versus Passive

In the conventional sense all helicases are active enzymes. That is to say that they use the energy derived from the hydrolysis of ATP to drive movement. However, the term active or passive has been coined to describe enzymes that either bind at a double stranded region

and separate the duplex prior to unwinding (active), or enzymes that bind the ends of unravelled single strands, brought about either by an additional enzyme process or by thermal "fraying", before proceeding to unwind the DNA duplex (passive) (Lohman *et al.* 1996).

1.3 Classification of Helicases

Based on primary structure analysis, Gorbalenya and Koonin (1993) designated three superfamilies and two families of alleged helicases. The largest two superfamilies, superfamily 1 (SF1) and superfamily 2 (SF2) had core domains made up from seven conserved motifs; motif I (Walker A), Ia, II (Walker B), III, IV, V, and VI (Figure 1.1). However despite providing a fundamental framework for structure-function analysis, recent studies have exposed a number of insufficiencies as well as revealing additional motifs. Furthermore, it has become clear that the original core motifs are not exclusively indicative of helicases but rather of translocases; enzymes that couple ATP hydrolyis with directional movement along ribonucleic acids. Although helicases themselves are a subgroup of translocases, many proteins may not withstand their classification as helicases due to a lack of authentic helicase activity; the physical separation of duplex ribonucleic acids. Thus rendering the Gorbalenya and Koonin mode of classification unable to differentiate between helicases and translocases (Stanley *et al.* 2006).

Despite the possibility for considerable confusion there are some regular structural and mechanical themes, such as the translocase core domains; the minimal structural unit formed from neighbouring recombination protein (Rec) A-like domains. It is these domains that undergo conformational change upon ATP hydrolysis, providing a mechanical output from a chemical input (Ye *et al.* 2004). Two features of the core domains are the presence of conserved residues involved in binding and hydrolysis of ATP; this is comparable to the Walker A and B motifs found in many ATPases (Walker *et al.* 1982), and an 'arginine finger' utilised in coupling NTP binding and hydrolysis to protein conformational change by detecting the γ-phosphorylation state of the bound nucleotide (Scheffzek *et al.* 1997; Caruthers *et al.* 2002).

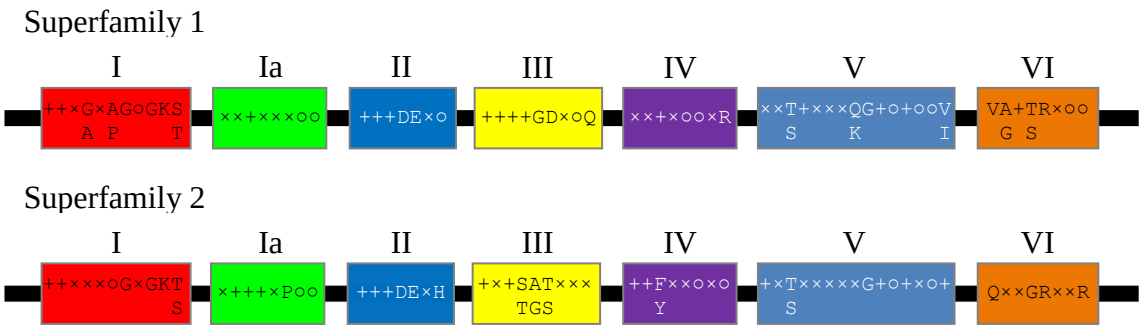


Figure 1.1 **SF1 and SF2 Motifs Adapted from Hall and Matson (1999).** Presented with the N-terminus on the left and C-terminus on the right, the coloured boxes represent the conserved helicase motifs and are relevant only to the motif's continuing colour theme throughout this text. The letters inside the boxes are those motif's consensus amino acid sequences, and the labels above are the motif's assigned names. The relative positions of motifs and spacing between them are not representative of actual proteins. In all consensus sequences + and ○ represent non-specific hydrophobic or hydrophilic residues respectively, while × represents residues not constrained to being hydrophobic or hydrophilic.

In response to all of the above, a more adequate form of classification has been proposed (Singleton *et al.* 2007). Consequently, helicases have been subdivided into six superfamilies; SF1, SF2, SF3, SF4, SF5, and SF6. The seven original motifs of SF1 and SF2 have been expanded on to contain the Q Motif (Tanner *et al.* 2003) and the TPGR motif (Pause *et al.* 1992) also referred to as motif Ib (Cordin *et al.* 2006), as well as the SF1 specific motif IVa (Korolev *et al.* 1998) and the SF2 specific TRG motif (Mahdi *et al.* 2003) (Figure 1.2). In addition, nomenclature based on translocation polarity (A for 3' to 5' and B for 5' to 3') and core domain substrate specificity (α for single-stranded nucleic acids and β for double-stranded nucleic acids) has also been introduced.

It should be noted that alternative nomenclature has been proposed by Fairman-Williams (2010) to include an alternative motif Ib, a conserved arginine or glycine residue in SF1 or SF2 respectively, while the TPGR motif is referred to as motif Ic. Furthermore, motif IVa in SF1 has been designated motif IIIa, motif IVa has been designated to the SF2 exclusive TRG motif, and additional motifs Va and Vb have been introduced. Acknowledging these differences, the nomenclature in this text will follow that described by Gorbalenya and Koonin (1993) and reviewed by Hall and Matson (1999), with the additional expanded motif nomenclature based on the original identifying studies.

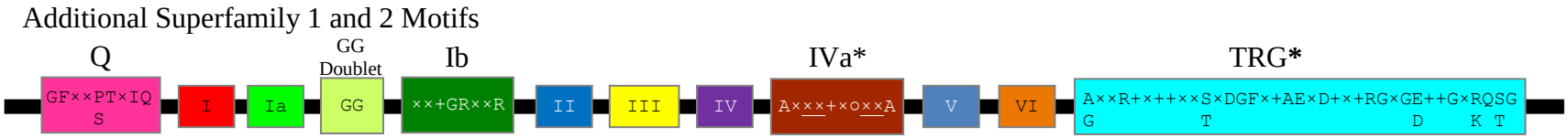


Figure 1.2 **The Additional SF1 and SF2 Motifs.** Presented with the N-terminus on the left and C-terminus on the right, the smaller boxes represent the conserved motifs presented by Gorbalenya and Koonin (1993) and described in Figure 1.1, while the larger boxes represent the motifs included by Singleton *et al.* (2007). The letters inside the larger boxes are those motif's consensus amino acid sequences, and labelling above the boxes are the motif's assigned names; Q, Ib, IVa, and TRG (Translocation in RecG), by Tanner *et al.* (2003), Pause *et al.* (1992), Korolev *et al.* (1998), and Mahdi *et al.* (2003) respectively. The GG doublet which was first identified by Linder *et al.* (1989) has been included for the sake of completeness, despite not being largely acknowledged as a motif. The presence of a * indicates that motif's identification in a SF1 species while a ***** indicates that motif's identification in a SF2 species. The colours of the boxes are relevant only to the motif's continuing colour theme throughout this text. The relative positions of motifs and spacing between them are not representative of actual proteins. In all consensus sequences + and ○ represent non-specific hydrophobic or hydrophilic residues respectively, while × represents residues not constrained to being hydrophobic or hydrophilic. Underlined characters are non-specific residues, where any or all may not be present in any particular species.

1.3.1 Superfamily 1 (SF1)

From a structural perspective SF1 helicases are the most comprehensively characterised. There has been considerable debate over SF1 oligomeric state, generally however they are considered monomeric or dimeric. All members appointed thus far are α enzymes but there is no consensus on translocation polarity with both A and B within the family.

1.3.1.1 SF1A

Two of the best studied members of the SF1A subfamily are the gram-negative bacteria helicase Rab escort protein (Rep) (Subramanya *et al.* 1996) and the gram-positive bacteria helicase plasmid copy reduced A (PcrA) protein (Korolev *et al.* 1997). The crystal structures of these proteins have revealed two RecA-like domains; termed the N-core and C-core (Subramanya *et al.* 1996) relating to the N and C-terminal domains respectively. These domains also referred to as 1A and 2A, each contain an additional subdomain insert; 1B and 2B respectively. Within the cleft between the N- and C-core domains there is an ATP-binding site lined with many of the SF1 motifs (Figure 1.3).

Crystallography studies of Rep in complex with ssDNA (Korolev *et al.* 1997) and PcrA in complex with dsDNA, as well as an non-hydrolysable ATP analogue and a sulphate ion (Velankar *et al.* 1999), reveal similar conformational changes. Combined with biochemical data such as the kinetic parameters of ATP hydrolysis (Soultanas *et al.* 1999; Soultanas *et al.* 2000) and a step size of one base pair per ATP (Dillingham *et al.* 2000), a model for duplex unwinding could be proposed. This, although based on studies of Rep and PcrA, provided a canonical model for all SF1 helicases (Soultanas *et al.* 2000; Soultanas *et al.* 2001) (Figure 1.4).

Interaction between the helicase and a single-stranded nucleic acid results in the movement of subdomain 2B, exposing a double-stranded nucleic acid interaction surface between itself and subdomain 1B. Then ATP binding induces the closure of the cleft between the 1A and 2A subdomains around the double-stranded nucleic acid, which in turn induces two further changes: Firstly, the movement of 1B and 2B which forces the duplex DNA onto a negatively charged surface, consequently destabilising the component strands; and

secondly, the flipping of the nucleotide bases within the nascent single-stranded region between acceptor pockets in the 1A and 2A domain cleft. The movement of these bases, which gives rise to translocation along ssDNA, combined with the binding, hydrolysis, and release of ATP, resembles Yarranton and Gefters' (1979) proposed 'inchworm' method of DNA unwinding.

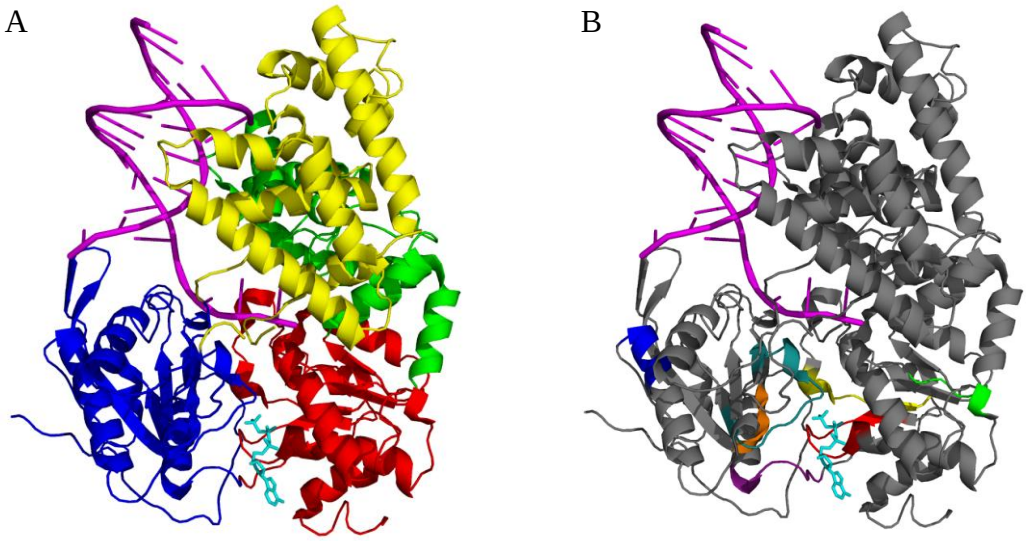


Figure 1.3 **Crystal Structure of PcrA Bound to dsDNA and ATP (Velankar *et al.* 1999).** Rendered using PyMol v0.99 PDB code 3PJR; the assumed biological molecule of PcrA. Both renders A and B show the bound dsDNA in magenta (top left) and the ATP ligand (bottom centre) in cyan. Render A shows the N and C-core domains (1A and 2A) in the bottom right (red) and left (blue) respectively. The corresponding B subdomains are shown in the upper half, with subdomain 2B coming out of the plane (yellow), and the subdomain 1B going into the plane (green). Render B shows the arrangement of the conserved motifs (Hall *et al.* 1999) surrounding the ATP binding cleft; motif I (red), motif Ia (green), motif II (blue), motif III (yellow), motif IV (deep purple), motif V (deep teal), and motif VI (orange).

1.3.1.2 SF1B

Typical examples of the SF1B helicase subfamily are RecD and DNA-dependant ATPase (Dda), although work has been performed on other lesser studied SF1B members such as phytochrome interacting factor 1 (Pif1) and RNA recognition motif 3 (Rrm3) (Ivessa *et al.* 2002). RecD is a subunit of the RecBCD complex (Exonuclease V) which initiates recombinational repair (Kowalczykowski 2000) and provides the only structural insight

into the 1B subfamily. Some biochemical studies have been performed but are in the context of the RecBCD complex. Contrastingly, Dda is well biochemically studied but lacks structural studies. As a result theories presenting mechanical details have been drawn from a combination of these works.

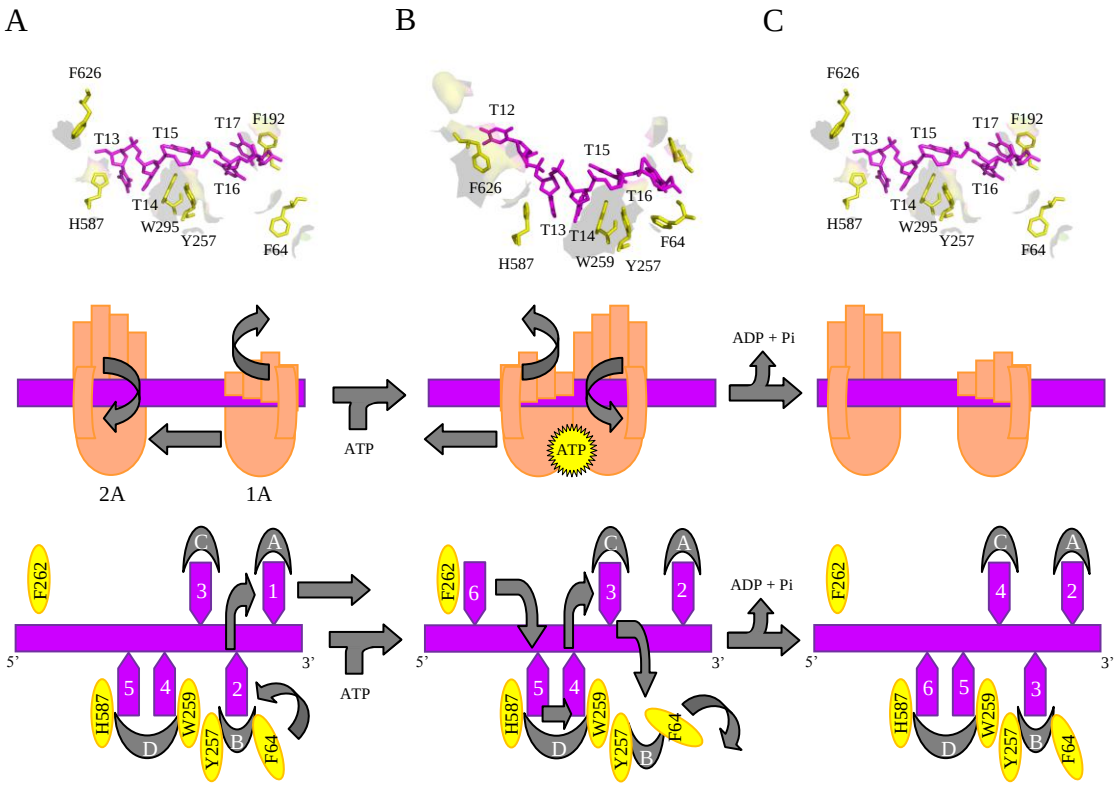


Figure 1.4 **The Translocation Mechanism of SF1 Enzymes (Adapted from Singleton *et al.* (2007)).** Based on the intermediate crystal structures of Rep (Korolev *et al.* 1997) and PcrA (Subramanya *et al.* 1996) with bound ssDNA during alternative ATP hydrolysis states. A to C show the three translocation steps of PcrA associated with ATP-binding and hydrolysis. The top row shows the stacking between the amino acid residues (yellow) and the ssDNA bases (magenta). The middle row is an analogy of each stage; the gripping hands depict which domains have tight or loose grips on the bound ssDNA. The bottom row shows a cartoon of the structural data in the top row, with the bases (1-6) switching between the enzymes nucleotide binding pockets labelled A-D.

Both RecD and Dda exhibit monomeric activity (Taylor *et al.* 1995; Nanduri *et al.* 2002), albeit in the case of RecD as part of a heterotrimer containing an SF1A helicase (RecB) (Korangy *et al.* 1993). RecD structure reveals two RecA-like core domains either side of an ATP binding site, echoing the SF1A helicases. However, there is also a further N-terminal domain that interfaces with RecC. This is not a conserved feature within Dda. An

observable difference that this may bring is that RecD does not interact with duplex DNA ahead of the fork within the RecBCD complex. dsDNA interaction however cannot be attributed to the Dda N-terminal or any other SF1B members based on a single observation. It is more likely that the mechanism of translocation follows the inchworm method of the core domains similar to that of the SF1A subfamily.

1.3.1.3 The Active Form of SF1 Helicases

Deliberation over oligomeric state has led to the conclusion that the most likely form of SF1 helicases at the fork is monomeric. This was based on the observation that all structural studies present the enzymes to be monomeric under a range of crystallisation conditions (Subramanya *et al.* 1996; Korolev *et al.* 1997; Velankar *et al.* 1999), and models based on this data involve helicase monomers at the fork (Velankar *et al.* 1999; Soultanas *et al.* 2000). Dda single-turnover experiments have shown that monomers of the enzyme are capable of unwinding dsDNA (Ivessa *et al.* 2002). In contrast to this wild-type Rep helicase requires multiple monomers to unwind the duplex (Ha *et al.* 2002). However, monomers of the truncated form which lacks the 2B domain can actively unwind independently suggesting a non-standard regulatory role of the domain in Rep (Brendza *et al.* 2005).

PcrA and ultraviolet (UV) resistant D (UvrD) monomers have been shown to, if nothing more, translocate along ssDNA (Dillingham *et al.* 2000; Fischer *et al.* 2004). It should however be acknowledged that, although the mechanism of cooperation is not fully understood the presence of multiple monomers does enhance the efficiency of helicase activity *in vitro*. Multiple UvrD monomers have been shown to be more effective at unwinding dsDNA flanked by longer ssDNA tails (Ali *et al.* 1999). Likewise, multiple Dda monomers can cooperatively unwind DNA substrates (Byrd *et al.* 2005), as well as displace streptavidin from biotinylated DNA more effectively than independent monomers (Byrd *et al.* 2004).

1.3.2 Superfamily 2 (SF2)

SF2 is the largest superfamily of helicases and members have been shown to translocate on either single-stranded or intact duplex nucleic acids (α and β respectively). It contains several comprehensively studied subfamilies such as the DEAD-box RNA family (Cordin *et al.* 2006), the RecQ-like family (Bennett *et al.* 2004), and the sucrose non-fermenting 2 (Snf2)-like family (Flaus *et al.* 2004; Flaus *et al.* 2006). Like SF1 helicases their oligomeric state is considered to be either monomeric or dimeric. Some SF2 members have been shown to operate as NTP-dependant non-processive mechanical switches that alter nucleic acid and nucleoprotein conformation (Linder *et al.* 2006).

1.3.2.1 Translocation on Single-Stranded Nucleic Acid

Hepatitis C virus (HCV) non-structural protein 3 (NS3) is a SF2α helicase. It has been shown to unwind both duplex RNA and DNA (Pang *et al.* 2002) although it does require a 3' single-strand overhang (Tai *et al.* 1996). Despite observed functionality as a homodimer (Levin *et al.* 2004; Tackett *et al.* 2005), active unwinding from monomeric NS3 has been reported (Dumont *et al.* 2006). NS3 structural studies have elucidated crystal structures unbound and bound to ssDNA (Kim *et al.* 1998) (Figure 1.5). N- and C-core domains have annotated functions in viral polyprotein processing and protein-protein interactions respectively (Mackintosh *et al.* 2006).

Reminiscent of SF1 helicases, the enzyme core contains two RecA-like domains. However, where the aromatic residues within the ssDNA-binding site of SF1 (exampled by PcrA) stabilise the elongated conformation of the ssDNA molecule by stacking against the bases, in SF2 the ssDNA maintains its shape as if in a B-form duplex, preserving stacking between the bases. There is an unconserved domain within NS3 which prevents (by steric hindrance) this pseudo B-form ssDNA from base pairing with a second ssDNA molecule. If it were possible for DNA base pairing to take place, the B-form duplex would be suitable for translocation (β activity) without prior unwinding. On the surfaces of the ssDNA-binding site are residues from motifs Ia, Ib, IV, and V that directly contact the nucleic acid, principally involving the phosphodiester backbone. In addition, a valine residue at position 432 and a tryptophan residue at 501 form two 'bookends' that group

five bases, while the regions of the RecA-like domains surrounding these bookends form two structurally equivalent DNA contact regions.

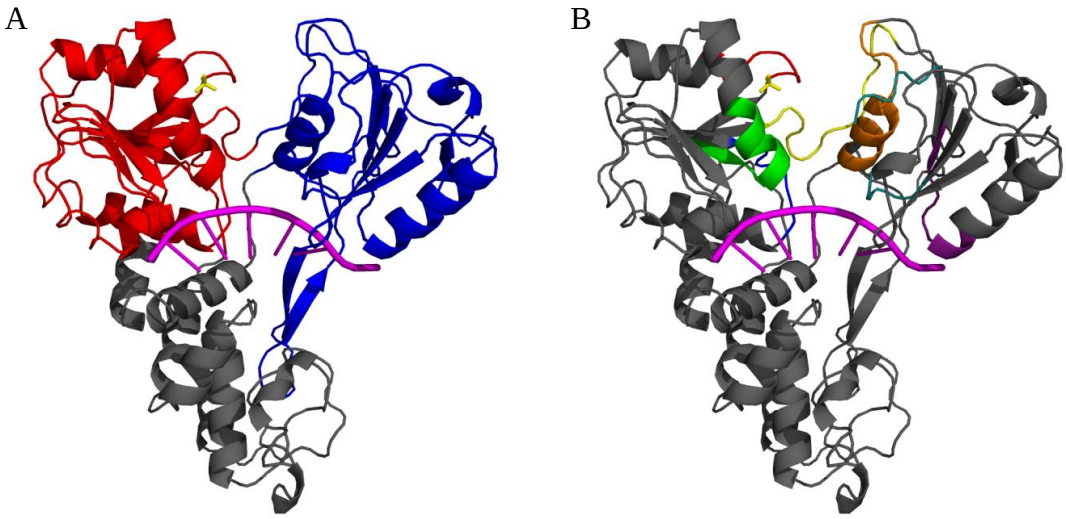


Figure 1.5 **Crystal Structure of HCV NS3 Bound to ssDNA and a Sulphate Ion (Kim *et al.* 1998).** Rendered using PyMol v0.99 PDB code 1A1V; the protease domain is omitted. Both renders A and B show the bound ssDNA in magenta and the sulphate ion, marking the ATP-binding site, is shown as a stick-model at the top (yellow). Render A shows the N- and C-core domains in the top left (red) and right (blue) respectively. Render B shows the arrangement of the conserved motifs (Hall *et al.* 1999) surrounding the ATP binding cleft; motif I (red), motif Ia (green), motif II (blue), motif III (yellow), motif IV (deep purple), motif V (deep teal), and motif VI (orange).

Mutagenesis (Paolini *et al.* 2000; Tai *et al.* 2001; Kim *et al.* 2003) directed towards these residues which lie outside of the fundamental motifs has led to the proposal that they play a part in the ratchet translocation mechanism (Kim *et al.* 1998). In response to ATP hydrolysis and product release the core domains open and close altering the distance between the bookends. If this is coordinated with the binding and release of the DNA substrate it would result in enzyme translocation and possibly DNA unwinding. Information on ATP binding induced changes has been determined by studying another SF2 RNA helicase; DEAD box polypeptide 4 (DDX4 previously named VASA).

Structural studies providing DDX4 bound to a non-hydrolysable analogue (Sengoku *et al.* 2006) have shown that NTP binding at motif I produces a closed conformation as residues, including the arginine finger of motif VI, close around the nucleotide. This mechanism is

similar to that of SF1 concluded from PcrA studies. However, there is one potential difference; the protein precisely binds to the substrate despite lacking the mechanism of un-pairing the duplex DNA by subdomains 1B and 2B. This cannot be regarded as fully accurate though, as present crystal derived structures have not contained duplex DNA. The lack of a single protein open and closed crystal structure also makes it difficult to assess step size. Presumably it would not be more than a few base pairs per hydrolysed ATP, but there is no reason why it should be as low as the PcrA step size (1bp/ATP).

Like SF1 helicases, SF2 also show evidence for potential cooperation between monomers. NS3 studies have shown large step sizes, although these observations may be due to the cooperative effect of multiple smaller unwinding steps (Levin *et al.* 2004; Serebrov *et al.* 2004; Dumont *et al.* 2006).

1.3.2.2 Translocation on Double-Stranded Nucleic Acid

While there are comprehensive examples of a SF2α helicase, SF2β helicases are much less represented in the literature (Fairman *et al.* 2004; Stanley *et al.* 2006). This is mostly due to the difficulty in DNA tracking which produces no products. Triplex displacement assays have however been employed in several studies relating to translocase activity (Firman *et al.* 2000; McClelland *et al.* 2005).

The only SF2β crystal structure with duplex bound DNA across the terminus domains is from *Sulfolbus solfataricus* alpha thalassemia/mental retardation syndrome X-linked (ATRX previously named Rad54) protein (Durr *et al.* 2005) from the Snf2 family (Flaus *et al.* 2006) (Figure 1.6). This shows the repeated theme of two RecA-like domains, but with two inserts of unknown function. However, the C-core domain is rotated 180° from its positions in SF1 and SF2α helicases. This difference, although not identical, has been observed in other SF2 helicases (Caruthers *et al.* 2000; Sengoku *et al.* 2006). The inserts may prove to be related to conformational change resulting from ATP or substrate binding, not dissimilar to those reported in members of the DExD/H-box RNA helicases responsible for local RNA remodelling (Sengoku *et al.* 2006).

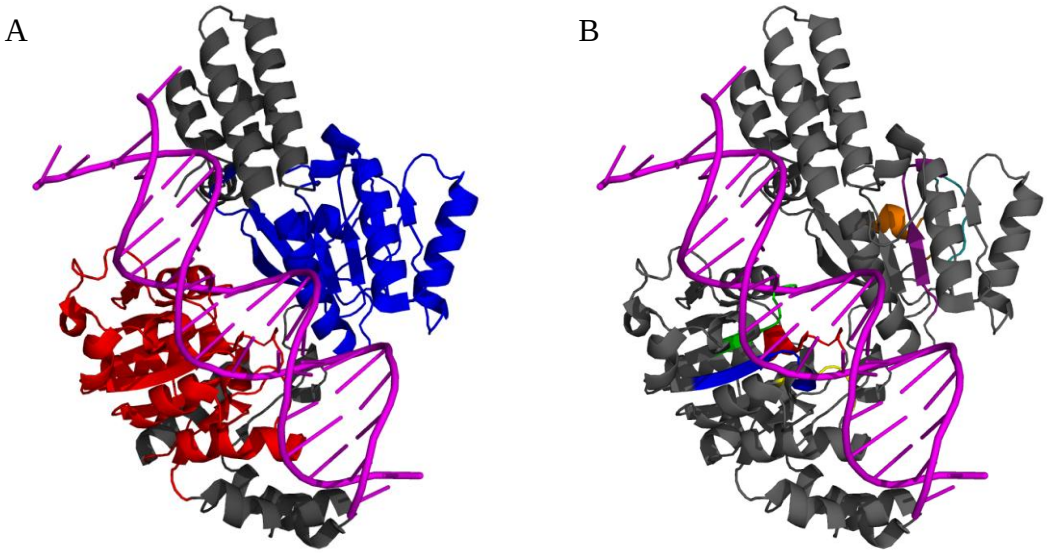


Figure 1.6 **Crystal Structure of Rad54 Bound to dsDNA (Durr *et al.* 2005).** Rendered using PyMol v0.99 PDB code 1Z63; the assumed biological molecule with the intact dsDNA ligand. Both renders A and B show the bound dsDNA in magenta. Render A shows the N- and C-core domains to the left (red) and right (blue) respectively. Render B shows most of conserved motifs (Hall *et al.* 1999) arranged within the N-core domain; motif I (red), motif Ia (green), motif II (blue), motif III (yellow), motif IV (deep purple), motif V (deep teal), and motif VI (orange).

The ATRX N-core domain structure in *Danio rerio* shows a standard open configuration which supports the paradigm that the domain must be open for translocation to progress (Thoma *et al.* 2005). The N-core domain of ATRX interacts with both DNA strands through contacts between its motif Ia and Snf2 specific motifs, and the minor grooves of the phosphodiester backbones. During binding, the DNA remains largely in the B-form. Substrate interaction with the C-core domain is minimal, likely due to its orientation. The positioning of one of the component single-strands of the duplex relative to the N-core reflects that of NS3; it also contains the majority of DNA-protein contacts. These observations suggest that ATRX follows an inchworm translocation mechanism, as described for PcrA and NS3, capable of twisting the duplex DNA across the core domains but does not result in unwinding.

When observing studies of EcoR124I translocase activity of modified or discontinued substrates, changes on the 3’ to 5’ strand showed greater levels of inhibited translocation than that of changes on the 5’ to 3’ strand (Stanley *et al.* 2006). This supports a model suggesting a principal motor interaction with the 3’ to 5’ component strand.

1.3.2.3 Polarity of SF2 Helicases

While the majority of SF2 helicases have type A polarity, type B exceptions have been identified. An example of type B are the Xeroderma Pigmentosum D (XPD) family, which contribute to the polymeric human transcription factor II complex (TFIIH), and are characterised by the presence of an iron-sulphur cluster in the N-terminal domain. Polarity for double-stranded substrate translocation depends on enzyme loading so can be difficult to define, although polarity experiments utilising modifications to either of the component strands of the duplex can give some idea (Stanley *et al.* 2006). Results from these studies support the observation presented earlier, demonstrating a particular importance of one of the two strands, imparting a unifying translocation mechanism across the superfamilies 1 and 2. Cordin *et al.* (2006) discussed several non-processive DEAD-box helicases with bidirectional activity. Rather than being translocases, it may be that these enzymes act as remodelling switches (Linder *et al.* 2006). Given the compelling structural similarities across the SF2 helicases, it is likely that the conformational switch mechanisms bear resemblance to a single directional step of processive motors. Supporting this is the observation that these enzymes do display processivity in response to protein partner interactions.

1.3.2.4 Accessory Domains: Targeting SF2 Enzymes

The RNA helicase eIF4A is an example of a SF2 helicase where the active polypeptide chain is located exclusively in the core domains. The majority of SF2 helicases however, include accessory domains within either or both of the core domains or N and C-flanking regions. The domains seem to impart specificity or complementary catalytic activity. Protein partners may act to regulate helicase activity, or to confer accessory functions. This is supported by the observation that helicases often contribute to hetero polymeric complexes.

The core domains of several SF2 helicases have been shown to be functionally modified by targeting specific nucleic acid structures. RecG, a SF2Aβ helicase, converts replication

forks into Holliday junctions. Crystallography of RecG in complex with a replication fork analogue (Singleton *et al.* 2001) has revealed three domains. The first includes a 'wedge' that directs specificity to the fork junction, while the second and third form the core motor domains (Figure 1.7). The wedge guides the template strand toward the core motor in a similar direction to that of the SF2α ATRX-dsDNA complex.

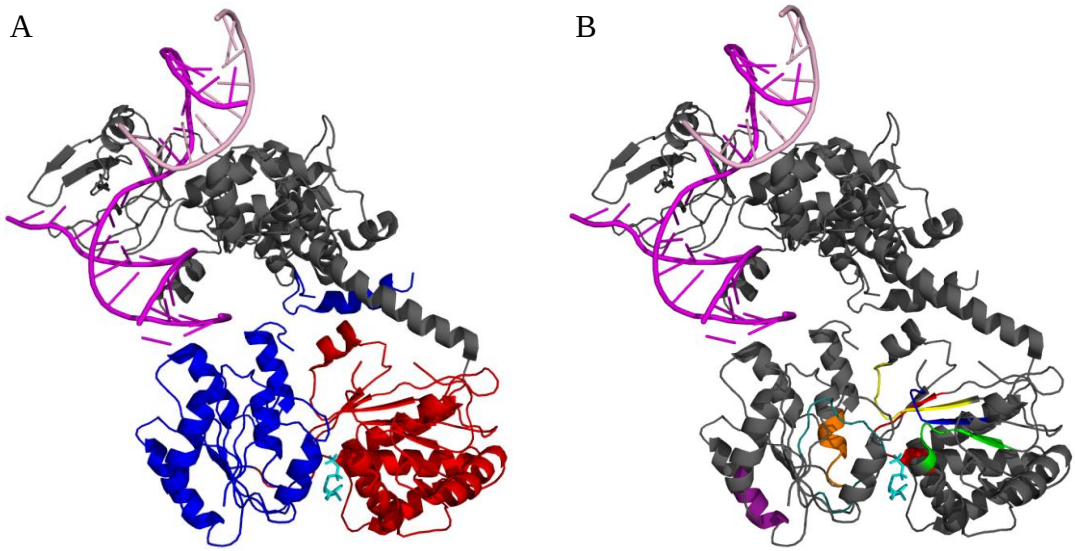


Figure 1.7 **Crystal Structure of RecG Bound at a Three-Way DNA Junction with ATP (Singleton *et al.* 2001).** Rendered using PyMol v0.99 PDB code 1GM5; demonstrates how the N-terminal accessory wedge (grey) acts as a specificity factor for the three-way DNA junction. Render A shows the DNA duplex (magenta) coordinated towards the N- and C-core domains shown in red and blue respectively, in a similar fashion to that of Rad54 (Figure 1.5). The lagging nascent strand is represented in light pink. The ATP (cyan) binding site is shown situated in the cleft between the core domains (bottom). Render B shows most of conserved motifs arranged within the N-core domain; motif I (red), motif Ia (green), motif II (blue), motif III (yellow), motif IV (deep purple), motif V (deep teal), and motif VI (orange).

Another nucleotide targeting helicase is UvrB, a subunit of the UvrABC Endonuclease which locates damaged DNA in the bacterial Nucleotide Excision Repair (NER) pathway. A β-hairpin extends from the N-core domain between the component strands of the duplex DNA (Linder *et al.* 2006). The hairpin provides specificity for damaged DNA, and functions as a stable clamp at sites of damaged DNA, which then recruits UvrC nuclease to excise the damage. Other examples of helicases containing accessory domains that could

be involved in targeting substrates include the RecQ-like enzymes which contain multiple DNA-binding accessory domains (Bennett *et al.* 2004).

Accessory domains have also been shown to target RNA structures, such as the RNA helicase DbpA, which contains a modular domain which confers specificity for a particular hairpin within the 23S ribosomal RNA (rRNA) of the large prokaryotic ribosomal subunit (Diges *et al.* 2001).

Other helicase targets include nucleoprotein complexes. Remodelling and disruption of various DNA-protein interactions by Snf2 family members have been demonstrated (Flaus *et al.* 2006). As well as containing a nucleic acid-binding core SF2β motor, many Snf2 members may interact directly with target proteins (Grune *et al.* 2003). The action of both of these interactions simultaneously, *id est* DNA translocation to or from a bound target protein, could result in nucleoprotein remodelling (Flaus *et al.* 2004; Durr *et al.* 2005), and the generated torque may prove an important role in remodelling mechanisms.

Unfortunately, information required for a full understanding of the relative core motor and protein interaction domain geometries of any Snf2 family members is not available. However, alternative enzyme structure studies might contribute to the hypothesis. For instance, the bacterial transcription-coupled repair factor Mutation Frequency Decrease (Mfd), which displaces stalled RNA polymerase from damaged template DNA, reveals a SF2 motor and RNA polymerase interaction module (Deaconescu *et al.* 2006). Other examples are the SF2 DExD/H-box RNA helicases, where studies indicate that they can effectively catalyse protein displacement from RNA, thereby participating in the structural remodelling of ribonucleoprotein assemblies (Jankowsky *et al.* 2001) such as the spliceosome (Linder *et al.* 2001).

The addition of complementary catalytic domains can produce novel enzymatic properties in helicases. For example, the juxtaposition of helicase and nuclease domains is common in SF2 systems; Werner (WRN), type I restriction endonucleases, and helicase-associated endonuclease for fork-structured DNA (HEF) (Komori *et al.* 2002). Another example is the reverse gyrase of thermophilic bacteria, which combines a SF2 helicase domain with a type IA topoisomerase domain, and is capable of catalysing the ATP-dependent positive supercoiling of DNA.

1.3.4 Hexameric Helicases

Gorbalenya and Koonin (1993) originally described one hexameric superfamily, SF3, and two hexameric families, replicative DNA helicase B (DnaB)-like, and rhodopsin (Rho)-like. Under the Singleton *et al.* (2007) mode of characterisation SF3 has retained its status as a superfamily, DnaB- and Rho-like have been promoted to SF4 and SF5 respectively. Furthermore, they are joined by the ATPases Associated with various cellular Activities (AAA^+) proteins (Erzberger *et al.* 2006), which have been designated as SF6. All four hexameric superfamilies show considerable structural similarity. They are all toroid, consisting of six monomers containing either a RecA or an AAA^+-like ATPase core. The sites of nucleotide-interaction are localised at the surfaces that make up the monomer interfaces, and the nucleotide binding pockets usually contain an arginine finger which is contributed to by neighbouring subunits and is involved in NTP binding and hydrolysis. Based on their stark similarity, it could be suggested that the four hexameric superfamilies should be combined and condensed into fewer superfamilies, although the system proposed by Singleton *et al.* (2007) refrained from doing so to avoid creating confusion around existing literature.

Hexameric helicases are best represented unwinding at the DNA replication fork. As a result, these helicases have to be highly processive as they are required to unwind from a few kilobases to several megabases, and in these circumstances processivity can be increased in a variety of ways. For example, several studies have shown the resolution of specific DNA topologies by various helicases (Webb *et al.* 2007; Long *et al.* 2009). Linking topology with helicase activity, notably its initiation, greatly increases processivity, possibly because topological features are more pronounced and are therefore more likely to result in an encounter between enzyme and substrate.

The closed ringed structure of hexameric helicases, and placement of the nucleotide-binding pockets on the internal ring surface, means that although some can self-assemble around the DNA substrate without additional cofactors, others require that they are loaded onto the substrate DNA. Thus, many hexameric helicases have ATP-dependant loading partners (Davey *et al.* 2003). The difference between the two adopted strategies is likely a

control mechanism for highly processive helicases, which if unchecked could lead to detrimental activity within the cell.

1.3.4.1 Superfamily 3 (SF3)

SF3 consists exclusively of helicases encoded by small DNA and RNA viruses, or by their prophage contribution to their host genomes (Iyer *et al.* 2004). They are associated with multiple enzymatic activities, most notably origin recognition and unwinding (Hickman *et al.* 2005). They can form double-hexamers and have a 3'-5' translocation polarity (type A). Originally only three conserved motifs were identified and all were located within a ~100 amino acid region. These were identified as A, B, and C. However, a fourth motif has been identified and coined B' due to its location between B and C (Koonin 1993) (Figure 1.8). Motifs A and B resemble the Walker A and B motifs similar to that of SF1 and SF2 helicases, B' contains a looped structure that interacts with the phosphodiester backbone of a bound nucleic acid molecule, while C is a SF3 specific motif and contains the Arginine finger.

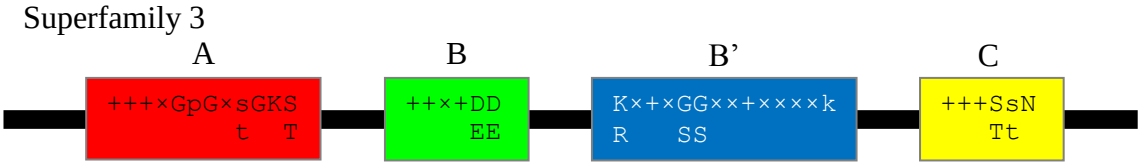


Figure 1.8 **SF3 Motifs Adapted from Hall (1999) and Koonin (1993).** Presented with the N-terminus on the left and C-terminus on the right, the coloured boxes represent the conserved helicase motifs. The letters inside the boxes are those motif's consensus amino acid sequences, and labelling above the boxes are the motif's assigned names. The colours of the boxes are relevant only to the motifs continuing colour theme throughout this text. The relative positions of motifs and spacing between them are not representative of actual proteins. In all consensus sequences uppercase characters represent highly conserved residues while lowercase characters represent lesser conserved residues. + and × represent non-specific hydrophobic or residues not constrained to being hydrophobic or hydrophilic respectively.

Several SF3 helicases have had their crystal structure solved (Figure 1.9) such as the simian virus 40 (SV40) (Li *et al.* 2003), adeno-associated virus (AAV) type 2 (James *et al.*

2003; James *et al.* 2004), and human papillomavirus serotype 18 (HPV18) (Abbate *et al.* 2004). All three examples are domains of larger initiator proteins; however AAV is also expressed individually via an alternative promoter. These structural studies have shown a common trait in the helicase domains relative position, in that it is always upstream of an origin-binding domain. Furthermore, each of the origin-binding domains possesses the same overall fold despite a low sequence similarity (Luo *et al.* 1996; Enemark *et al.* 2000; Hickman *et al.* 2002).

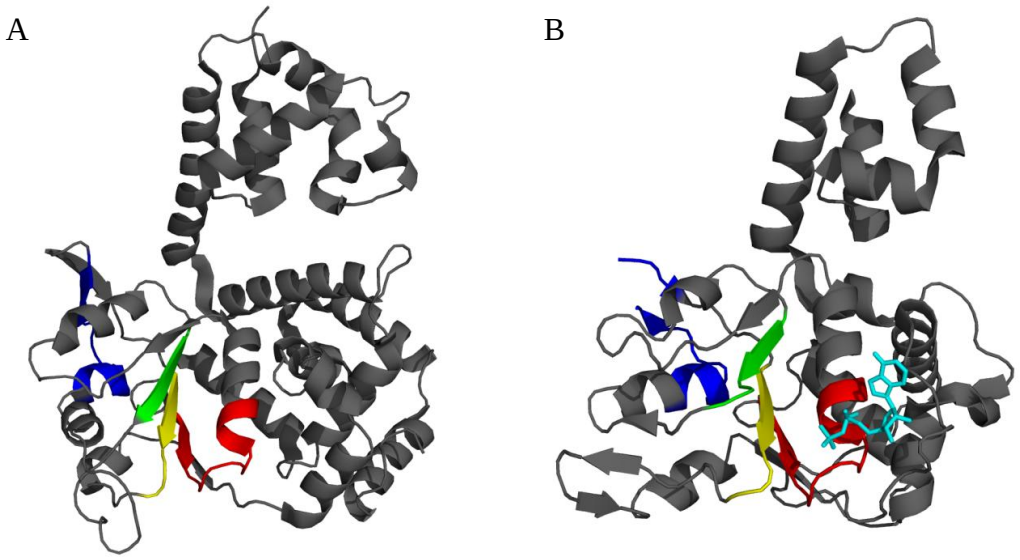


Figure 1.9 **Crystal Structures of SF3 Helicases SV40 (Li *et al.* 2003) and AAV2 with Bound ADP (James *et al.* 2004).** Rendered using PyMol v0.99 PDB codes 1N25 (SV40) and 1U0J (AAV2); Illustrates the relative positioning if the conserved motifs and the N and C-terminal regions. In both renders A and B the SF3 motifs A, B, B', and C are shown red, green, blue, and yellow respectively. AAV2 [B] shows a bound ADP (cyan) indicating the ATP binding site.

In contrast to the SF1 and SF2 helicase domains consisting of RecA-like fold, the common feature of the SF3 helicase domain is the presence of a modified AAA^+ core. The two folds are related (Koonin 1993), and like the RecA-like fold, the AAA^+ contains the nucleotide binding motifs within the central β-strands and loop inserts, functioning either in DNA binding or in protein-protein interaction. An additional common feature of SF3 helicases is a small poorly conserved helicase domain upstream of the AAA^+ core which may play a role in hexamer stabilisation.

1.3.4.2 Superfamily 4 (SF4)

SF4 helicases were original identified from bacterial and bacteriophage replicative systems. In bacterial systems the helicases exist as a separate polypeptide to the primase, such as the DnaB helicase and the DnaG primase of *E. coli*. However, many bacteriophage systems code for a single polypeptide which features both helicase and primase activity (Ilyina *et al.* 1992). All SF4 helicases thus far characterised display type B (5'-3') polatity.

Ilyina *et al.* (1992) described five conserved SF4 motifs; H1, H1a, H2, H3, and H4 (Figure 1.10). Motifs H1 and H2 are comparable to the Walker A and B motifs respectively. Motifs H1a, H3 and H4 bear no equivalents in any other helicase family; however H3 does contain a functional glutamine, occasionally substituted for histidine, which is functionally equivalent to motif 3 of SF1 helicases. Unlike the DEAD/H-box containing motif 2 (Walker B) of SF1 and SF2 helicases, H2 contains an isolated apsartate (D) among a hydrophobic region. Additionally, in SF4 helicases the glutamate (E) is replaced by a spatially equivalent residue from H1a.

Figure 1.10 **SF4 Schematic Diagram Adapted from Hall and Matson (1999).** Presented with the N-terminus on the left and C-terminus on the right, the coloured boxes represent the conserved helicase motifs. The letters inside the boxes are those motif's consensus amino acid sequences, and labelling above the boxes are the motif's assigned names. The colours of the boxes are relevant only to the motifs continuing colour theme throughout this text. The relative positions of motifs and spacing between them are not representative of actual proteins. In all consensus sequences + and ○ represent non-specific hydrophobic or hydrophilic residues respectively, while × represents residues not constrained to being hydrophobic or hydrophilic.

The best studied example of SF4 helicase is T7 bacteriophage gene 4 protein (gp4). The enzyme is composed of both a primase and a helicase domain separated with a flexible linker (Toth *et al.* 2003). Crystal structures elucidated a RecA-type core fold (Sawaya *et al.* 1999) and also provided some insight to nucleotide binding (Singleton *et al.* 2000) (Figure

1.11). Mutational analysis suggests that H4 projects a loop into the central channel of the hexamer that plays a role in ssDNA contact (Washington *et al.* 1996). While some SF4 studies support the suggestion of symmetric hexamers (Niedenzu *et al.* 2001), gp4 studies show deviation from the assumed six-fold symmetry with DNA-binding loops and arginine fingers occupying alternative conformations in different subunits. Additional observations of both symmetric and asymmetric conformations have been observed and it is has been proposed that this heterogeneity is representative of the enzyme's catalytic state (Barcena *et al.* 2001).

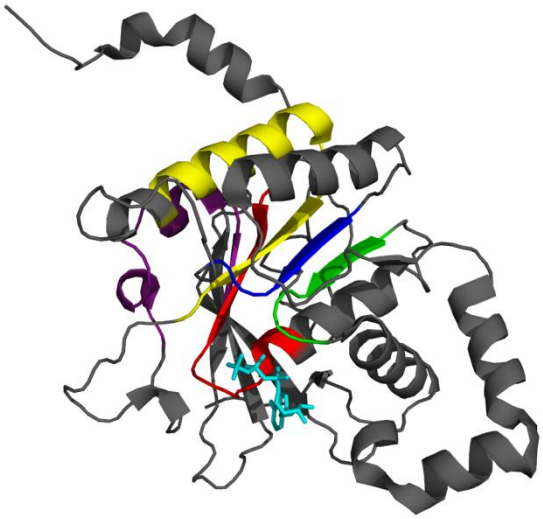


Figure 1.11 **Crystal Structure of a pg4 Monomer Bound to ADPNP (Singleton *et al.* 2000).** Rendered using PyMol v0.99 PDB code 1E0J; the assumed biological molecule of pg4. The ATP binding site is shown with the bound ADPNP in cyan, the conserved SF4 motifs H1, H1a, H2, H3, and H4 are shown in red, green, blue, yellow, and deep purple respectively.

1.3.4.3 Superfamily 5 (SF5)

Despite being closely related to SF4, Rho-like helicases have been designated SF5 based on their sequence. Rho binds to a specific sequence on the nascent RNA strand and unwinds it from the template DNA strand during transcription termination in bacteria.

Electron microscopy studies of Rho have revealed both open and closed forms of the protein, where the open form has been interpreted as being either incomplete (lacking a

monomer) or an open toroid (Gogol *et al.* 1991). The solved crystal structure (Skordalakes *et al.* 2003) also shows the open form. It is likely that this structure represents a loading intermediate as it displays adjacent monomers at the break rising away from each other providing a gap sufficient enough for the passage of ssRNA into the central channel.

1.3.4.4 Superfamily 6 (SF6)

A number of AAA+ core containing proteins are not classified as SF3 helicases and so have been appointed SF6 helicases by Singleton *et al.* (2007). One instance is the mini chromosome maintenance (MCM) protein complex which consists of six monomers (MCM2-7). It is thought to be the main replicative helicase of eukaryotic organisms and has shown to be vital for replication initiation and elongation (Labib *et al.* 2000), although helicase activity has only been detected in a heterohexamer of MCM4, 6, and 7 (Kaplan *et al.* 2004). Another SF6 example is prokaryotic resistance to UV (Ruv) B, which is involved in homologous recombination along with RuvA and RuvC in the processing of Holliday junctions (West 1996).

No crystal structures for eukaryotic MCM have been reported, and inferences about its structure have largely been drawn from homohexameric archaeal MCM homologues (Chong *et al.* 2000). An N-terminal crystal structure (Fletcher *et al.* 2003) and electron microscopy studies (Pape *et al.* 2003) of archaeal MCM show a large central channel through the hexameric ring that is sufficiently wide enough to accommodate dsDNA.

Monomeric crystal structures of RuvB with a bound nucleotide have been solved (Putnam *et al.* 2001; Yamada *et al.* 2001). Comparisons to other AAA+ protein crystal structures without bound nucleotide suggest a conformational change that might be coupled to translocation activity. Superimposition of RuvB to electron microscopic data, combined with the MCM data, provides a consistent model which couples the conserved arginine fingers with nucleotide hydrolysis, and presents a central channel which accommodates dsDNA.

The MCM and RuvB AAA+ proteins were chosen only for example purposes, more extensive and representational reviews of AAA+ studies are available from Erzberger and Berger (2006), and Iyer *et al.* (2004).

1.3.4.5 Mechanisms of ATP Hydrolysis

A common feature of hexameric helicases is their multiple NTP-binding sites exhibiting strong heterogeneous cooperation (Bujalowski *et al.* 1993). Debate surrounds the exact details of nucleotide-binding and hydrolysis and as a result multiple models have been proposed (Figure 1.12).

The three-site sequential model (Hingorani *et al.* 1997) involves three of the six NTP-binding sites equally distributed around the hexamer. Each binding site binds an NTP, hydrolyses it into P_i and NDP, and then releases the NDP leaving the site empty. These configurations are designated T (triphosphate), D (diphosphate), and E (empty) respectively, and no two sites can be in the same configuration.

The all sites sequential model (Singleton *et al.* 2000) is similar to the three-site sequential model, differing only in that all six NTP-binding sites are occupied meaning that any two will be in the same configuration. A variation of this model includes a blocking system by neighbouring sites in the same configuration (Enemark *et al.* 2006).

The stochastic model (Martin *et al.* 2005) details all six of the NTP-binding site as catalytically active but each site can be in any of the three configurations at any given time. Finally, like the stochastic model, the concerted model (Gai *et al.* 2004) describes all the NTP-binding sites as catalytically active but they are all are synchronised in their configuration. It is likely that more than one of the proposed models is correct and relates to the particular species being described.

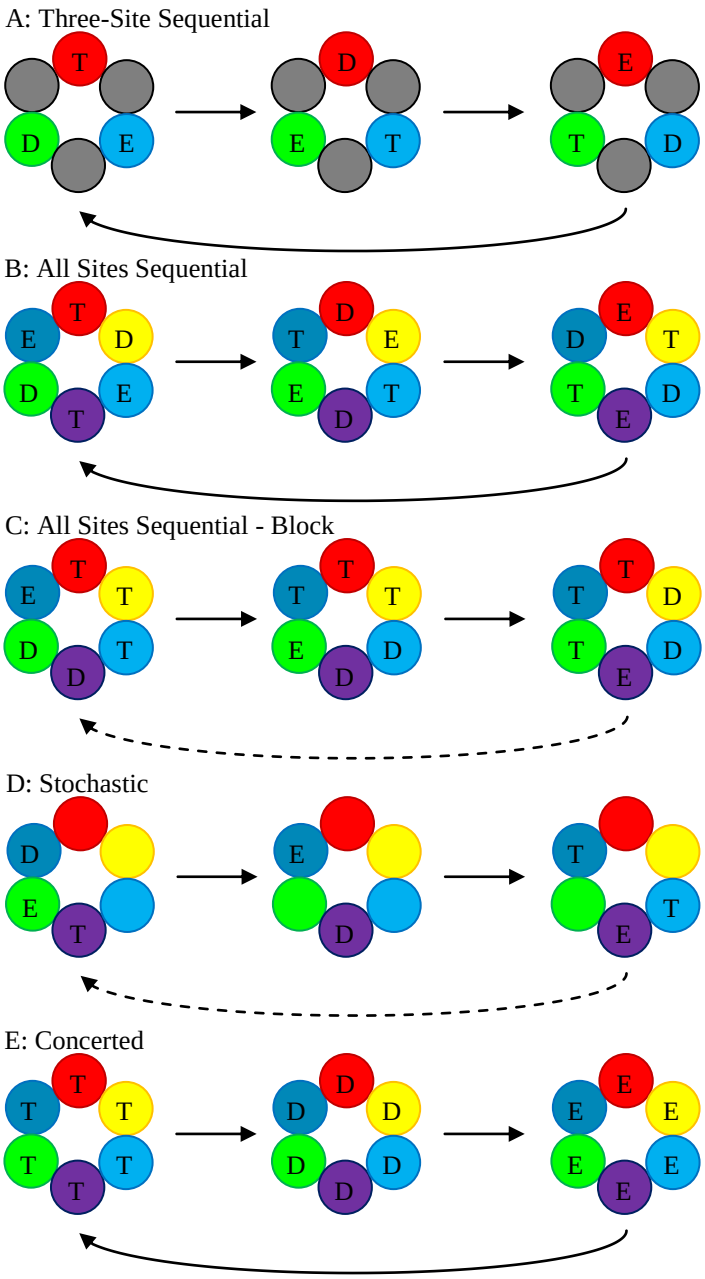


Figure 1.12 **Models for Hexameric Helicase NTP Hydrolysis and Release (Singleton *et al.* 2007).** Schematics A-E detail the various models for NTP hydrolysis; they are the three-site sequential model [A] (Hingorani *et al.* 1997), the all sites sequential model [B] (Singleton *et al.* 2000), the all sites sequential – block model [C] (Enemark *et al.* 2006), the stochastic model [D] (Martin *et al.* 2005), and the concerted model [E] (Gai *et al.* 2004). The coloured circles resemble individual monomers of the toroid ring, their colours only serve to show that each monomer remains in its relative position and that only it hydrolysis state alters, this is indicated by the letters T (NTP bound), D (NDP bound), and E (empty). Solid arrows linking the last illustrated stage with the first [A, B, and E], show that there are only three stages of the cycle. Broken arrows indicate that there are either an indeterminate number of steps [D] or that there are more than three steps [C (has 6 steps)].

1.3.4.6 Coupling NTPase Activity to Translocation

All of the proposed NTP-hydrolysis models are coupled to translocation of the helicase along DNA. Hexameric helicases feature looped projections into the toroid's central channel. Mutational studies of these loops has shown them to be essential for DNA binding (Washington *et al.* 1996). Structural analyses (Singleton *et al.* 2000; Enemark *et al.* 2006) suggest the loops track the phosphodiester backbone with loop movement driven by NTP-binding and hydrolysis, however the precise character of the mechanism which drives loop movement and controls DNA-binding affinity has not been determined and will probably differ between different helicases. Similarly, step size for hexameric helicases has not been explicitly determined and although examples do exist (Kim *et al.* 2002; Galletto *et al.* 2004), this too will likely be dependent on the individual helicase.

1.4 Helicases as Antiviral Drug Targets

Despite considerable progress in the development of antivirals over the last 15 years (De Clercq 2004), there is still a call for new drugs to improve existing treatments and to combat viral resistance. Although many viral infections can be successfully prevented via vaccinations and so this will likely remain the staple for the treatment of these infections, there is little effective knowledge of vaccines for many important human pathogens. This has resulted in the proposal of alternative strategies, one of which is the emergence of helicases as potential novel targets for the antiviral treatments.

Viruses replicate either by using self-encoded proteins or by hijacking their host cellular proteins (Kwong *et al.* 2005). This presents two distinct strategies which in theory could target either a viral or cellular polypeptide with equal effectiveness (De Clercq 2002), although each strategy presents alternative repercussions. Targeting viral helicases decreases the chance of creating inhibitors that are toxic to the host but also increases the requirement for specificity as the viral and cellular helicases catalyse similar enzymatic reactions. Nevertheless, viral and cellular helicases are not identical and medicinal chemistry and structure-based drug design can take advantage of these differences in designing drugs which are highly specific to viral proteins.

Under the present strategy, resistance will occur over time unless an antiviral drug is extremely potent, so that it rapidly reduces the pool of viral replication, leading to shorter treatment time, preventing the proliferation of resistant viruses (Richman 1996). Targeting key proteins required for viral replication should eliminate the problem of resistance. Hypothetically, a pan-antiviral drug could be designed to inhibit all viruses that depend upon a common host factor; however a likely disadvantage would be an increased risk of toxicity to the host (Kwong *et al.* 2005). In practice, drug design is influenced by whether the virus' replication is performed autonomously using self-encoded genes, or whether the virus recruits the host protein.

1.4.1 *Plasmodium falciparum*

There are approximately 257 species of parasitic protozoan in the *Plasmodium* genus (WHO 2008). *Plasmodium falciparum* is one of at least 10 human malaria causing plasmodium strains, and is responsible for over 90 % of worldwide cases of human malaria and has the highest reported complications and mortality rates when compared to any other human malaria-causing *Plasmodium* (WHO 2009; WHO 2010). Unlike other *Plasmodium* species, *P. falciparum* is restricted to warmer climates due to its temperature dependant development (Figure 1.13).

Malaria infection develops in two stages; the exoerythrocytic which takes place outside the red blood cell (RBC), and erythrocytic which takes place inside the RBC. When an infected female A*nopheles gambiae* mosquito takes a blood meal from a host animal, sporozoites enter the host bloodstream, migrate to the liver, and infect the host hepatocytes. Here they multiply asexually and asymptomatically for between eight and thirty days (Bledsoe 2005). Many species of *plasmodium* have long been known to display a dormant stage at this point, and more recently numerous cases of dormancy in *P. falciparum* have been reported (Greenwood *et al.* 2008; Szmitko *et al.* 2009; Theunissen *et al.* 2009).

After liver-stage development, the parasite differentiates into thousands of merozoites. The parasite induces the death and detachment of the hosts' hepatocytes releasing the merozoite-filled vesicles, called merosomes. These rupture and discharge the merozoites into the host bloodstream where they infect RBCs (Sturm *et al.* 2006; Tarun *et al.* 2006)

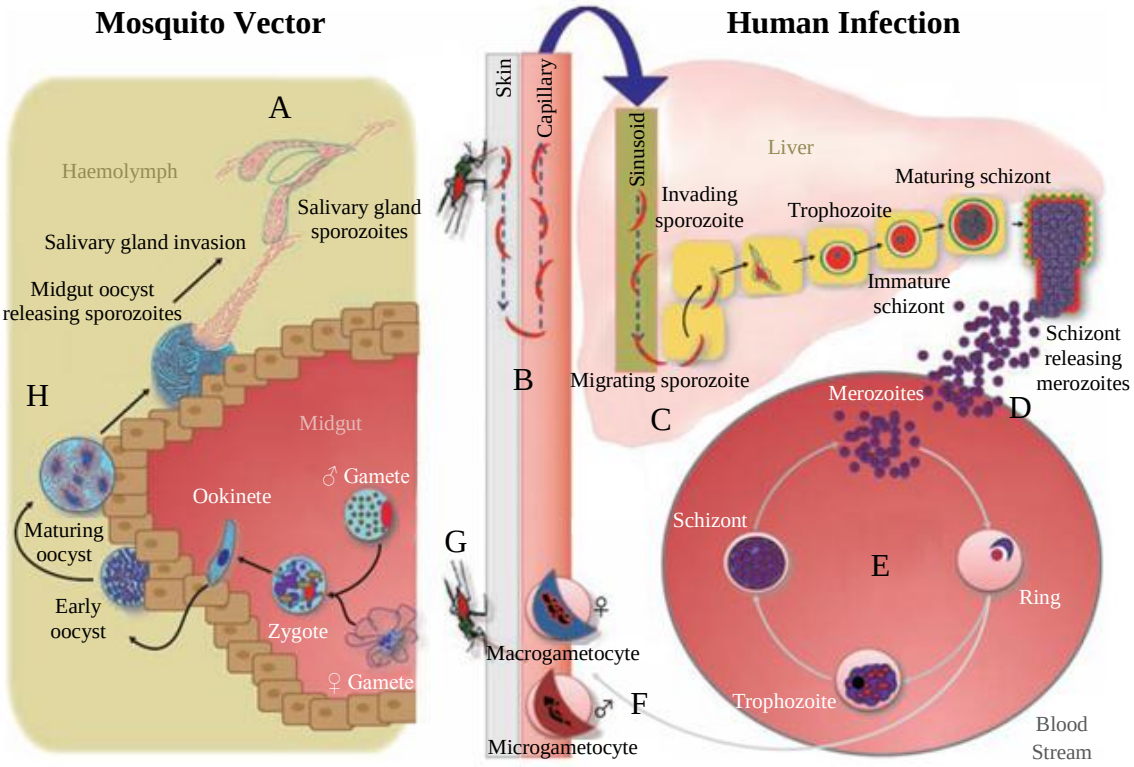


Figure 1.13 **The *Plasmodium* Life Cycle Adapted from Kappe (2010).** Starting at mosquito salivary gland infection [A], the motile sporozoites are injected into human skin during a mosquito feed where they migrate into the tissue and invade a blood vessel [B]. Sporozoites are transported, via the bloodstream into the hepatocytes of the liver through the sinusoid vessels [C], initiating liver-stage development. A single sporozoite undergoes massive growth to form tens of thousands of merozoites which are subsequently released back into the bloodstream. Each merozoite infects an erythrocyte [D], commencing the intraerythrocytic cycle comprised of an asexual blood stage replication [E] where repeated cycles of infection, replication, and release of next-generation merozoites results in the release of sexual-stage gametocytes [F]. The gametes are taken-up during a mosquito feed and mate to form a zygote in the mosquito midgut [G]. The zygote develops into a motile ookinete which penetrates the mosquito haemolymph where it develops into an oocyst [H]. Once fully mature the oocyst releases sporozoites which infect the mosquito salivary gland, thus beginning the cycle over.

beginning the erythrocytic stage. In the parasitised RBCs (pRBC) the parasite asexually multiplies further, the periodic release and re-infection into neighbouring RBCs is associated with the waves of fever described within the classical symptoms of malaria (Bledsoe 2005). The pRBC often undergoes membrane modification, forming protrusions giving a knobbed appearance. This results in the infected RBCs binding to other cell surfaces, such as other RBC and the cells which make up the inner surface of small blood vessels (Aley *et al.* 1984; Newton *et al.* 2000). This helps protect the parasite harbouring

cells from being circulated and destroyed in the spleen, but can also result in microvascular blockages causing placental and cerebral malaria (Miller *et al.* 1994). The pRBCs have also been shown to breach the blood-brain barrier in cerebral malaria potentially leading to patient coma (Adams *et al.* 2002).

Some merozoites differentiate into male and female gametocytes, which can be taken up during an *A. gambiae* mosquitoes' feed. In the mosquito mid-gut, the gametocytes mature and the male microgamete fertilises the female macrogamete, undergoing sexual recombination to form the zygotes (Kappe *et al.* 2010). The zygote develops into an ookinete which penetrates the lining of the mid-gut to form the oocyst with a new generation of sporozoites under the outer-gut lining. The sporozoites migrate to the mosquitos' salivary gland, completing the parasite life cycle.

After World War II vigorous efforts to eradicate malaria were put into practice (Litsios 1996). Despite success over large geographical regions, malaria endures in tropical Africa and much of Asia. Subsequent healthcare initiatives which focused responsibility for malarial control on peripheral healthcare workers were also largely unsuccessful. Interest in the control of malaria has become increasingly focused on the worsening situation in sub-Saharan Africa and its contributory factors (Greenwood *et al.* 2002) such as climate instability and global warming (Hay *et al.* 2002; Tanser *et al.* 2003), war and civil-unrest resulting in healthcare breakdown (Bloland *et al.* 2002), interaction with and increased susceptibility to malaria due to Human Immunodeficiency Virus (HIV) infection (Whitworth *et al.* 2000; French *et al.* 2001; Mount *et al.* 2004; Kublin *et al.* 2005), and travel between non-endemic and endemic areas. However, the main cause is likely the spread of drug-resistant parasites and insecticide-resistant *A. gambiae* mosquitoes (Snow *et al.* 2001; Korenromp *et al.* 2003).

1.4.1.1 Drug-Based Approaches to Combat Malaria

Despite being largely ineffective and having potentially underestimated side effects, such as retinal toxicity (Farrell 2012), chloroquine is still the present strategy against uncomplicated *P. falciparum* malaria in many countries. The consequences of drug-resistant parasites is such a problem that the World Health Organisation (WHO) advocates

a combination of vector control strategies such as insecticidal treated nets (ITNs) and indoor residual spraying (IRS), and parasitic interventions such as Intermittent Preventative Treatment in Pregnancy (IPTp) and Intermittent Preventative Treatment for Infants (IPTi) using sulphadoxine-pyrimethamine (SP) (WHO 2010). However, many countries that have substituted for SP have reported parasitic resistance (EANMAT 2003; Plowe *et al.* 2004).

Kremsner and Krishna (2004) called for a combination therapy approach to be adopted to increase efficacy and decrease treatment duration, but despite reported improvement some countries have yet to adopt this approach. Questions arise in respect to what combinations of anti-malarial drugs to use. Taking into account effectiveness and affordability, artemisinin-based combination treatments (ACTs) have been used with good results from both Asia (Mayxay *et al.* 2004) and regions of Africa (Mutabingwa *et al.* 2005; Piola *et al.* 2005). More recently however, resistance to artemisinin has been reported around the Thai-Cambodian border (Dondorp *et al.* 2009).

Approaches in optimising present treatment and devising new ones have been transformed in recent years due to more sophisticated laboratory techniques (Vernick *et al.* 2004) looking at more accurate identification of the infecting *Plasmodium* species (Singh *et al.* 2004), and focusing on genetic markers characterising the parasitic virulence (Jensen *et al.* 2004; Staalsoe *et al.* 2004) and drug-resistance (Sutherland *et al.* 2002; Djimdé *et al.* 2003). Studies based on these observations have shown that the relatively long half-lives of many drugs, chloroquine and SP for example, provide the parasites with a selective advantage (Hastings *et al.* 2002; Hallett *et al.* 2004) however it seems that these resistant parasites are not transmitted as well as wild-type parasites. This has led to a concept that the discontinuation of treatment with these drugs may lead to a decrease in parasitic resistance, increasing success rates upon reintroduction of these drugs in a combination treatment scheme (Mita *et al.* 2003).

1.4.1.2 Molecular Approaches to Combat Malaria

Over a century after Sir Ronald Ross received the 1901 Nobel Prize in medicine for demonstrating the role of mosquitoes in the transmission of malaria (Brey 2003), the

genome of *A. gambiae* has been sequenced (Holt *et al.* 2002). From this have come new opportunities for targeted measures of control. Hallem *et al.* (2004) propose the identification of ligand inhibitors to the *A. gambiae* olfactory system which can then be utilised in new insect traps and repellents. Other possibilities have emerged regarding the genetic control over the mosquito population by means of introducing a modified strain of *A. gambiae* recalcitrant to *Plasmodium* into the wild population (Coleman *et al.* 2004), or by introducing germline transformants with genetic modifications that impede the normal functioning of female mosquitoes (Alphey *et al.* 2002).

Molecular studies have identified innate mosquito immune response proteins that attack the plasmodium ookinetes and oocysts (Cirimotich *et al.* 2010) and could be used as anti-parasitic molecules. Additionally, studies on sympatric *A. gambiae* populations show reduced *P. falciparum* infection intensities compared to allopatric populations (Harris *et al.* 2012). Studies on bacteria harboured in the mosquito midgut (Cirimotich *et al.* 2010; Rodrigues *et al.* 2010) have revealed that they play a significant role in tempering *Plasmodium* transmission, providing an alternative line of study focused on blocking infection of the insect.

Alternative to molecular targets elucidated from *A. gambiae*, many studies have focused on *Plasmodium* proteins themselves. One such stand out study follows *P. falciparum* reticulocyte binding protein homologue 5 (PFRH5) (Crosnier *et al.* 2011). In order for the parasite to gain entry into its human host's RBCs, PFRH5 relies on the human glycoprotein blood system determinant, extracellular matrix metalloproteinase inducer (EMMPRIN). A study in animal models demonstrates a cross-strain vaccine induced an antibody response to the RH5 antigen, raising anti-PFRH5 immunoglobulin (IgG) which neutralised numerous strains of *P. falciparum* (Douglas et al. 2011).

A second example is the phase three trial of RTS,S/AS01 (Agnandji *et al.* 2011), a recombinant vaccine which induces an antibody response to a cell surface protein of the *P. falciparum* sporozoite, called circumsporozoite. A potential short-coming in these approaches has been highlighted in studies on the erythrocyte membrane modification protein *P. falciparum* Erythrocyte Membrane Protein 1 (PFEMP1). Expressed during the erythrocytic stage, is it synthesised and transported from the intracellular parasite to the surface of the pRBC leaving them exposed to the host immune system. Protruding

PFEMP1 proteins have been identified as the binding interface of pRBCs to the brain endothelium of patients who have died from severe malaria (Adams *et al.* 2002).

Obviously these make an attractive molecular target, however there are multiple varieties within a single parasite and even more within that parasite population (Maier *et al.* 2009). Moreover, as the host immune system responds to the single expressed PFEMP1, sub-populations within the infecting parasite pool can interchange between their variant PFEMP1 surface proteins, restoring the infection and therefore remaining recalcitrant to the host immune system or conventional vaccinations (Chen *et al.* 2000; Pasternak *et al.* 2009).

1.4.1.3 Targeting Helicases to Combat Malaria

Upon the completion of the *P. falciparum* genome in 2002 (Gardner *et al.*), a full set of helicases were identified and annotated (http://plasmodb.org/plasmo/). Soon after, Tuteja (2007) proposed the potential for *P. falciparum* helicases as novel targets for effective anti-malarial drugs. Detailed Bioinformatic studies revealed 22 full-length putative DEAD-box helicases, and some other non-DEAD-box putative helicases (Seow *et al.* 2005; Suntornthiticharoen *et al.* 2006; Tuteja *et al.* 2006; Tuteja 2010). These proteins contain the conserved helicase domains and motifs, but vary in the N- and C-terminal regions and their coding sequences (Tuteja *et al.* 2006). At present only a few of these helicases have undergone any biochemical characterisation (Gupta *et al.* 2006; Suntornthiticharoen *et al.* 2006; Pradhan *et al.* 2007; Pradhan *et al.* 2008; Shankar *et al.* 2008; Tuteja *et al.* 2009; Prakash *et al.* 2010; Mehta *et al.* 2011).

Included within these characterised proteins is DEAD-box helicase *Plasmodium falciparum* helicase 45 kDa (PFH45, PlasmoDB gene ID: PF14_0655), an eIF4A homologue (Pradhan *et al.* 2007). So far studies have suggested that PFH45 is expressed through all the developmental stages of the parasite, binds both DNA and RNA, unwinds in both the 3'-5' and 5'-3' direction, and parasitic development is inhibited upon interference at both the genetic and proteomic level (Pradhan *et al.* 2007; Tuteja 2010; Tuteja *et al.* 2010). Pradhan *et al.* (2008) also report an up-regulation of helicase activity when phosphorylated with protein kinase C (PKC) and an inhibition of resolving

replication fork-like substrates in the presence of DNA-interacting molecules. Truncated forms have shown ATPase activity to be localised in the N-terminus while DNA/RNA binding is localised in the C-terminus, and neither truncated form is capable of independent unwinding activity. Several protein partners have also been identified for PFH45 (Tuteja *et al.* 2010).

Initial characterisation has also been performed on the 60 kDa truncated version of *Plasmodium falciparum* DEAD-box helicase (PFDH60, PlasmoDB gene ID: PFL310c), which shares a high degree of identity with DDX5 (previously named p68) (Pradhan *et al.* 2005); a highly conserved helicase from yeast to human. Unlike PFH45, PFDH60 expression is mainly during the schizont stage of parasitic development (Pradhan *et al.* 2005). Also, it exhibits helicase activity across a broad pH range (pH 5.0-10.0), requires a ssDNA region (Pradhan *et al.* 2007), and has been shown to unwind dsDNA, dsRNA, and DNA-RNA hybrids (Pradhan *et al.* 2005). Otherwise, like PFH45, PFDH60 shows bidirectional unwinding and displays up-regulated activity upon PKC phosphorylation.

1.5 Aims and Objectives

The principal purpose of this study is to explore the biochemical and structural properties of helicases from *P. falciparum* with an aim to exploit them as putative drug-targets. This will first involve the production of novel recombinants of PFH45 and PFDH60 from clones provided by Dr Renu Tuteja (Malaria Group, International Centre for Genetic Engineering and Biotechnology, New Deli, India), the purification of the fusion-tagged proteins after optimisation of heterologous expression in *E. coli* to produce yields sufficient enough to investigate the storage of the homologous protein samples, and for their characterisation. This work is described in chapter three.

Biochemical characterisation is to be performed on both helicases. Initial *in silco* investigations into both PFH45 and PFDH60 will be undertaken to expand on and confirm previous observations. Then *in vitro* analysis on the helicase and ATPase activities of both PFH45 and PFDH60 will be performed to elucidate preferential substrates, conditions, and cofactors for unwinding. In addition, a long-standing collaboration between this laboratory and Prof. David Bensimon's laboratory (Ecole Normale Superieure (ENS), Paris, France)

offers the opportunity to perform single-molecule biochemical characterisation on both helicases. This work is described in chapter four.

Initial structural studies will be performed to elucidated information on both helicases. In-house nuclear magnetic resonance (NMR) and X-ray crystallography services at the University of Portsmouth (Hampshire, UK) as well as good connections with the Diamond Light Source (Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire, England) will be taken advantage of to assist in these studies. This work is described in chapter five.

Studies into novel approaches of helicase inhibition will be performed. The effect of a novel DNA binding drug, provided as part of a collaborative project with Prof. David Thurston (King's College London, London, UK) on the helicase activity of both helicases will be studied. Furthermore, aptamers will be raised against each helicase and their subsequent cloning and sequencing of any aptamers generated will be performed. These could then be used as a high-throughput generation of potential inhibitors to be investigated further. This work is described in chapter six.

The work carried out in this study is described in detail within each chapter, and the conclusions and potential for future studies are discussed in chapter seven.

CHAPTER TWO – METHODS AND MATERIALS

2.1 Electrophoresis Protocols

2.1.1 0.8 % TAE Agarose Gel

Materials

1. Agarose.
2. TAE buffer: 50 mM TRIS·HCl pH 8.0, 0.1 % glacial acetic acid, 1 mM EDTA, made up to the desired volume with distilled water and filter sterilised[1] immediately.
3. Ethidium bromide: 50 mg mL^{-1} in distilled water.
4. 5 × Agarose gel loading buffer: 50 % (v/v) Glycerol, 100 mM EDTA, 0.125 % w/v bromophenol blue, made up to the desired volume with distilled water.

Method

0.8 % agarose gels were prepared by adding 0.8 g of agarose to 100 mL of TAE buffer and boiling until fully dissolved. Once sufficiently cooled, ethidium bromine was added to a final concentration of 1 μg mL^{-1}, mixed by gentle agitation, and poured into the gel plate and allowed to set for approximately 30 minutes. The set gel was then submerged in the gel tank under TAE buffer. The samples were mixed with 5 × agarose gel loading buffer prior to loading alongside the appropriate HyperLadder DNA marker (Bioline) (Figure 2.1) and then resolved for 1 hour at 100 V. The resolved samples were visualised and photographed using an ultraviolet light box.

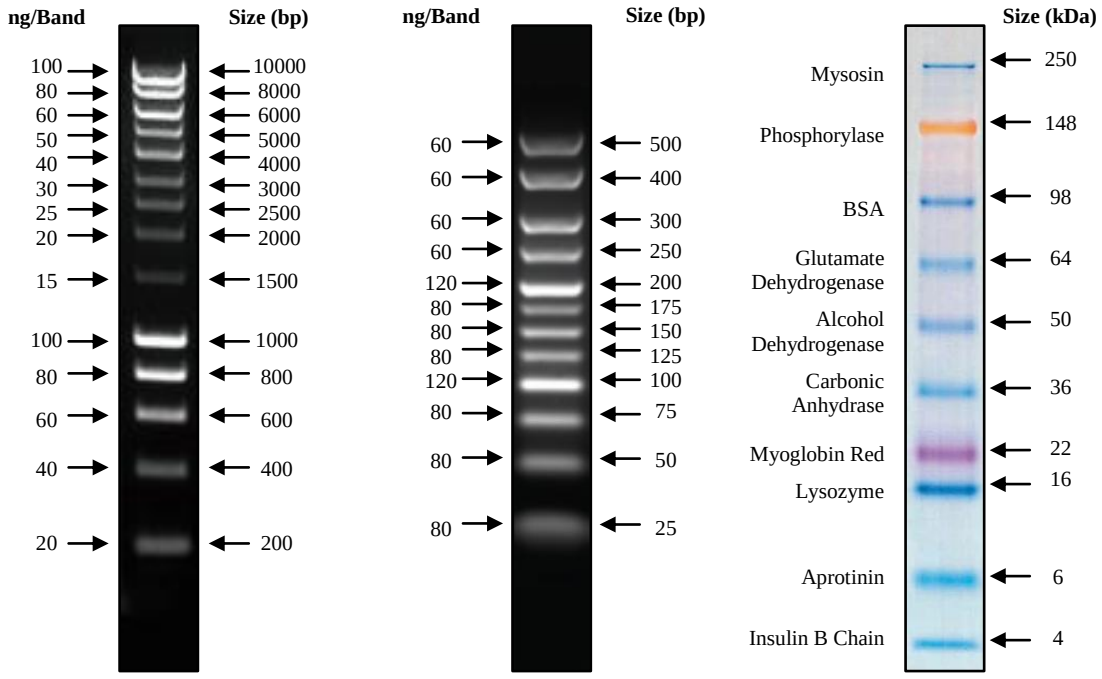


Figure 2.1 **Size Markers Used in this Study.** Adapted from the Bioline HyperLadder I product insert (left) and HyperLadder V (centre), 5 µL of HyperLadder I and V markers resolved using electrophoresis on a 1 and 3.5 % TAE agarose gel respectively, with ethidium bromide staining. The numbers to the left of the markers show the amount of DNA per band in nanograms (ng), and the numbers to the right show the corresponding size of each DNA band in base pairs (bp). Adapted from the Invitrogen SeeBlue Plus2 product insert (right), 10 µL of pre-stained SeeBlue Plus2 resolved using electrophoresis on a 10 % PAGE gel in tris-glycine running buffer. The labels to the left of the gel show the protein standards used in the marker, while the numbers to the right of the gel show the corresponding standard proteins size in kilodaltons (kDa).

2.1.2 10 % and 15 % SDS Polyacrylamide Gel Electrophoresis (PAGE)

Materials

1. SDS PAGE running buffer: 25 mM TRIS·HCl pH 8.0, 200 mM glycine, 2.5 mM sodium dodecyl sulphate (SDS), made up to the desired volume in distilled water.
2. Resolving gel solution: 10 % or 15 % Acrylamide (bis-acryl-amide (37:5:1)), 350 mM TRIS·HCl pH 8.0, 5 mM SDS, 0.1 % ammonium persulphate (APS), 0.1 % tetramethylethylenediamine (TEMED), made up to the desired volume using distilled water.

3. Stacking gel solution: 5 % Acrylamide (bis-acryl-amide (37:5:1)), 125 mM TRIS·HCl pH 6.8, 5 mM SDS, 0.1 % APS, 0.1 % TEMED, made up to the desired volume using distilled water.
4. 70 % Isopropanol.
5. 5 × SDS gel loading buffer: 50 mM TRIS·HCl pH 8.0, 100 mM EDTA, 20 mM SDS, 50 % (v/v) Glycerol, 0.125 % w/v bromophenol blue, 0.125 % w/v xylene cyanol, made up to the desired volume in distilled water.
6. Coomassie blue staining solution: 40 % Methoanol, 10 % ethanoic acid, 0.1 % w/v brilliant blue red-250 (BB·R-250), made up to the desired volume with distilled water.

Method

10 % and 15 % SDS PAGE gels were prepared by pouring 6 mL of resolving gel solution into the 8 mL PAGE gel cassette, overlaying with 70 % isopropanol and setting aside for 15 minutes. Once set, the isopropanol was decanted off and the resolving gel was rinsed with distilled water. 2 mL of stacking gel solution was then poured onto the set resolving gel. A well comb was inserted into the top of the cassette and set aside for another 15 minutes. The clamps and silicon gasket were then removed from the cassette, the cassette was placed into the gel tank, and the tank was filled with SDS PAGE running buffer. The well comb was removed from the cassette ready for the samples. The samples were mixed with 5 × SDS gel loading buffer and heated at 80 °C for 15 minutes prior to loading alongside a SeeBlue Plus2 Protein Marker (Invitrogen) (Figure 2.1). The gel was run for 20 minutes at 80 V to concentrate the samples in the stacking gel, and then 1 hour at 180 V to resolve the samples. The gel was stained with coomassie blue staining solution for 20 minutes at room temperature, and non-specific staining was removed by soaking in distilled water.

2.1.3 6 %, 10 %, and 15 % Native PAGE

Materials

1. TAE buffer: 50 mM TRIS·HCl pH 8.0, 0.1 % glacial acetic acid, 1 mM EDTA, made up to the desired volume with distilled water and filter sterilised[1] immediately.
2. Native gel solution: 6 %, 10 %, or 15 % Acrylamide (bis-acryl-amide (19:1)), 1 mM dithiothreitol (DTT), 0.1 % APS, 0.1 % TEMED, in TAE buffer.
3. 5 × Native gel loading buffer: 50 % (v/v) Glycerol.

Method

6 %, 10 %, and 15 % native PAGE gels were prepared by pouring 8 mL of native gel solution into the 8 mL PAGE gel cassette, inserting the well comb into the top of the cassette and setting aside for 15 minutes. The clamps and silicon gasket were then removed from cassette, the cassette was placed into the gel tank, and the tank was filled with TAE buffer. The well comb was removed from the cassette and the gel was pre-equilibrated for 1 hour at 100 V at 4 °C. The samples were mixed with 5 × native gel loading buffer prior to loading, the gel was run for 60 minutes at 100 V at 4 °C. The gel was visualised depending on the sample's specific requirements.

2.2 Purification Protocols

2.2.1 DNA Restriction Digest

Materials

1. DNA substrates; *PFH45* and *PfDH60* clones in pET-28a (Novagen) and pGEX-6P-1 (GE Healthcare) expression vectors (Figure 2.2).
2. Restriction enzymes (Table 2.1) (NEB).
3. 10 × Restriction digest buffer (NEB).
4. 10 × Bovine serum albumin (BSA) (NEB).
5. Distilled water.

Method

The restriction digests were carried in bullet tubes containing the DNA substrate with 1 unit (U) of enzyme per 1 µM of DNA divided by the number of minutes per reaction. The respective restriction digest buffer for the enzymes was added, as well as BSA to act as a crowding agent where needed. The reaction was made to a final volume using distilled water and incubated at 37 °C for 1 hour. Where applicable, the reactions were terminated by heat activation, with aliquots run on a 0.8 % agarose gel (2.1.1) to observe cleavage.

2.2.2 Preparation of Competent Cells (Hanahan *et al.* 1991)

Materials

1. Cell strains[2] (Table 2.2): *Escherichia coli* Dh5α (Invitrogen), *E. coli* BL21 (DE3) pLysS (Novagen).
2. SOB media: 2 % w/v Tryptone, 0.5 % w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, made with distilled water and autoclave sterilised[1] immediately.
3. CCMB80: 10 mM KOAc pH 7.0, 80 mM $CaCl_2 \cdot 6H_2O$, 20 mM $MnCl_2 \cdot 4H_2O$, 10 mM $MgCl_2 \cdot 6H_2O$, 10 % (v/v) Glycerol, made with distilled water, pH down to a minimum of 6.4 then autoclave sterilised[1] immediately.

Method

Incubation flasks containing 250 mL SOB media were spiked with 100 µL of stock cells and grown overnight at 20 °C. The following day, once an OD_{600} ~0.3 was reached, the cells were collected by centrifugation at 4000 × g at 4 °C for 12 minutes. The pellets were then re-suspended in 50 mL of 4 °C CCMB80 buffer and incubated on ice for 20 minutes. The cells were then centrifuged at 4000 × g at 4 °C for 12 minutes and re-suspended in 10 mL of 4 °C CCMB80 buffer. The cells were separated into 500 µL aliquots at stored at -70 °C.

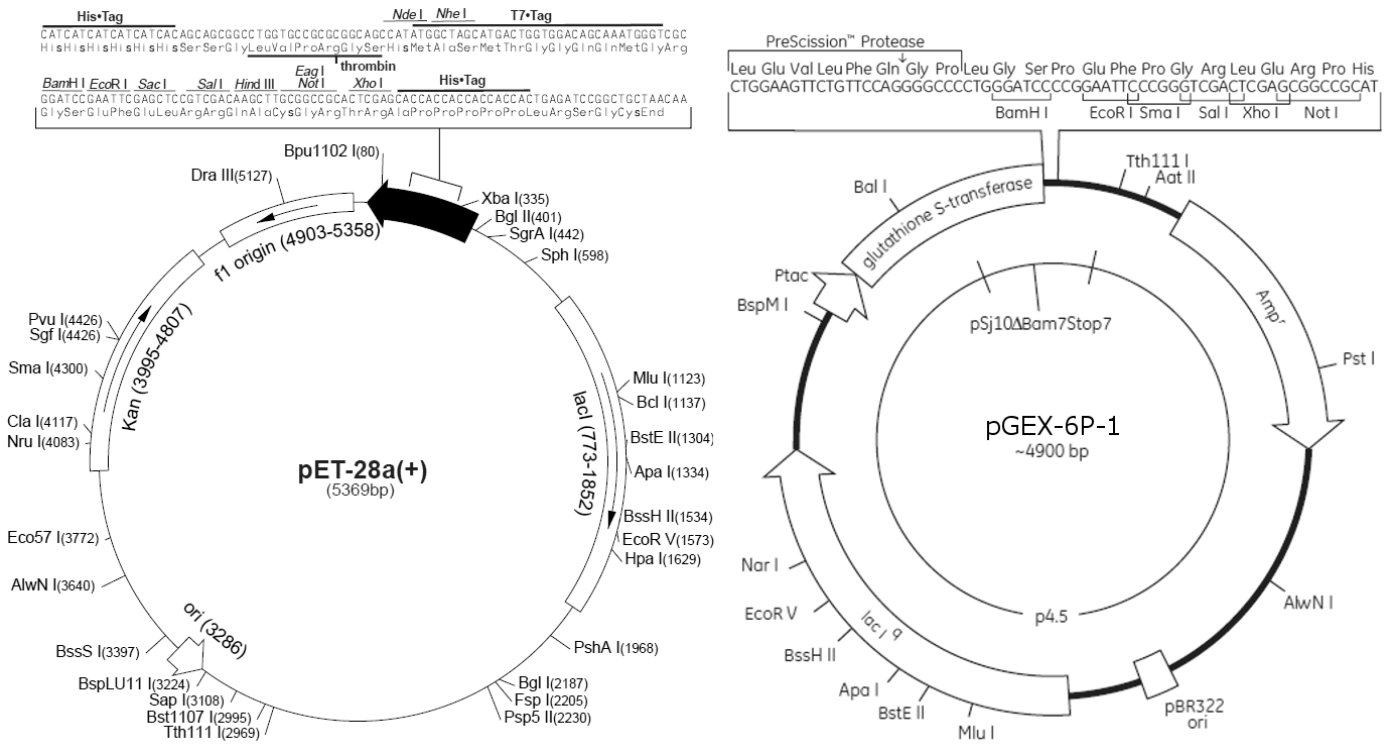


Figure 2.2 **Maps of the Expression Vectors used in this Study.** The pET-28a(+) map (left) has been adapted from the Novagen pET-28a(+) product insert and the pGEX-6P-1 map (right) is adapted from the GE Healthcare pGEX-6P-1 product insert. For each vector the multiple cloning site (MCS) is highlighted at the top, detailing the associated cleavage and restriction sites. The His-Tag coding region of pET-28a(+) is located with the MCS sequence and is under the control of a T7 *lac* promoter, the glutathione S-transferase (GST) coding region of pGEX-6P-1 is located upstream of the MCS and is under the control of a *tac* promoter.

Clone	Restriction Enzyme Upstream of Insert	Restriction Enzyme Downstream of Insert	NEBuffer for Double Digest	Optimum Working Temperature	BSA Requirement	Heat Inactivation
pET-28(+)/PFH45	BamHI	NotI	NEBuffer3	37°C	Yes	65°C 20 minutes (NotI)
Optimised pET-28(+)/PFH45	BamHI	NotI	NEBuffer3	37°C	Yes	65°C 20 minutes (EcoRI)
Optimised pGEX-6P-1/ PFH45	BamHI	EcoRI	NEBuffer EcoRI	37°C	Yes	65°C 20 minutes (NotI)
Optimised pET-28(+)/PFDH60	BamHI	NotI	NEBuffer3	37°C	Yes	65°C 20 minutes (EcoRI)

Table 2.1 **Restriction Digests Reagents for the Clones used in this study.** Identified from the NEB 2011-12 catalogue, and the Novagen and GE Healthcare support material for the pET-28a(+) and pGEX-6P-1 plasmid cloning vectors respectively. Table 2.1 shows the necessary restriction enzymes for cutting both upstream and downstream of the gene insert, the buffers compatible for the double digest, the optimum temperature for the digest, if the presence of Bovine Serum Albumin (BSA) is required or not, and whether the enzymes can be heat inactivated; the brackets indicate which, if any enzyme is affected.

Bacterial Strain	**Genotype**	**Antibiotic Resistance**	**Working Concentration**
E. coli Dh5α (Invitrogen)	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lac*ZΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–	None	Not applicable
E. coli BL21 (DE3) [pLysS] (Novagen)	F^- ompT gal dcm lon $hsdS_B(r_B^-$ $m_B^-)$ λ(DE3) pLysS(cm^R)	Chloramphenicol	36 µg mL^{-1} in 100% dimethylformaldehyde (DMF)
E. coli JM109 (DE3) (Promega)	*end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, λ–, Δ(*lac-pro*AB), [F′, *tra*D36, *pro*AB, $lacI^qZ\Delta M15$], lDE3	None	Not applicable

Plasmid	**Associated Affinity Tag**	**Tag Cleavage Recognition Protein**	**Antibiotic Resistance**	**Working Concentration**
pET-28a(+) (Novagen)	Polyhistidine (His_6)	Thrombin	Kanamycin	40 µg mL^{-1} in distilled water
pGEX-6P-1 (GE Healthcare)	Glutathione S-transferase (GST)	PreScission Protease	Ampicillin	100 µg mL^{-1} in distilled water
pGEM-T Easy (Promega)	None	Not applicable	Ampicillin	100 µg mL^{-1} in distilled water
pJS4M	None	Not applicable	Ampicillin	100 µg mL^{-1} in distilled water

Table 2.2 **Bacterial Strains and Plasmids used in this Study.** The top section of the table shows the bacterial strains used in this study, lists their genotype, any antibiotic resistance they have, and the working concentration of that antibiotic. The bottom section of the table shows the plasmid expression vectors used in this study, the associated affinity tag which they contribute to the fusion protein, the protein cleavage site to remove the tag from the expressed fusion protein, the antibiotic resistance they confer within the transformed cell, and the working concentration of that antibiotic.

2.2.3 Transformation

Materials

1. Competent cells: *E. coli* Dh5α (Invitrogen), *E. coli* Bl21 (DE3) pLysS (Novagen), *E. coli* JM109 (DE3) (Promega).
2. DNA: Original clones; pET-28a(+)/*PFH45*, pET-28a(+)/*PfDH60*. Codon-usage optimised clones (GeneArt); pET-28a(+)/*PFH45*, pET-28a(+)/*PfDH60*, pGEX-6P-1/*PFH45*, pGEX-6P-1/*PfDH60*. EcoR124I clones; pGEX-6P-1/*HsdR* and pJS4M.
3. Lysogeny Broth (LB) media mix (Fisher Scientific): 1.5 % w/v Tryptone, 0.5 % w/v yeast extract, 1 % w/v NaCl, made with distilled water and autoclave sterilised[1] immediately.
4. LB agar (Lennox L Agar) plates mix (Fisher Scientific): 1 % w/v Tryptone, 0.5 % w/v yeast extract, 5 % w/v NaCl, 1 % w/v agar, made with distilled water and autoclave sterilised[1] immediately. Once sufficiently cooled to touch, the appropriate antibiotic (Table 2.2) was added and the molten agar was poured into culture plates to a depth of approximately 5 mm. Once set the plates were stored at 4 °C until required for a maximum of two weeks.

Method

The transformations were performed in bullet tubes containing 100 µL of competent cells with 1 µL of DNA by incubating the mixture on ice for 20 minutes. The mix was then transferred to a heat block at 42 °C for 90 seconds followed by the addition of 1 mL of 37 °C pre-warmed LB media and incubated at 37 °C for 1 hour. The growth was then spread onto a LB agar plate containing the appropriate antibiotic and grown overnight at 37 °C.

2.2.4 DNA Extraction (Birnboim *et al.* 1979)

Materials

1. Lysogeny Broth (LB) media mix (Fisher Scientific): 1.5 % w/v Tryptone, 0.5 % w/v yeast extract, 1 % w/v NaCl, made with distilled water and autoclave sterilised[1] immediately.

2. Antibiotic (Table 2.2): Kanamycin (Fisher Scientific), ampicillin (Fisher Scientific), stock solutions were filter sterilised[1] immediately.
3. GTE lysis buffer: 50 mM Glucose, 25 mM TRIS·HCl pH 8.0, 10 mM EDTA, made with distilled water and autoclave sterilised[1] immediately.
4. NaOH-SDS: 0.2M NaOH, 1 % SDS, freshly prepared prior to use.
5. 7.5M Ammonium acetate, pH 7.8 without adjustment.
6. Isopropanol
7. 70 % Ethanol: Made with distilled water.
8. TE-RNase A: 10 mM TRIS·HCl pH 8.0, 1 mM EDTA pH 8.0, 20 % w/v RNase A.

Method

A 37 °C overnight incubation of 3 mL of LB media with the appropriate antibiotic was inoculated with a single plate colony of cells. The following morning 1 mL of the culture was centrifuged at 15000 × g for 5 minutes. The supernatant was discarded and the cell pellet was re-suspended in 200 µL of GTE lysis buffer. After 5 minutes at R/T, 400 µL of NaOH-SDS was added and mixed by inversion. After a further 5 minutes on ice, 300 µL ammonium acetate was added and mixed by inversion. After 10 minutes on ice the samples were centrifuged at 15000 × g for 5 minutes. The clear supernatant was transferred into a fresh bullet tube, 500 µL (0.6 vol.) of isopropanol was added then incubated at R/T for 10 minutes. The samples were then centrifuged at 15000 × g for 10 minutes and the supernatants discarded. Then 1 mL of 70 % ethanol was added to the pellet. The samples were centrifuged at 15000 × g for 5 minutes and the supernatants were discarded. The pellets were then dried by placing the tubes cap-open upside-down on tissue for 15 minutes at R/T. The pellets were dissolved in 100 µL TE-RNase and incubated on ice for 15 minutes. The isolated DNA was cleaned using a DNA spin column. The samples were visualised on a 0.8 % agarose gel (2.1.1).

2.2.5 Protein Expression Analysis

Materials

1. Lysogeny Broth (LB) media mix (Fisher Scientific): 1.5 % w/v Tryptone, 0.5 % w/v yeast extract, 1 % w/v NaCl, made with distilled water and autoclave sterilised[1] immediately.
2. Antibiotic (Table 2.2): Kanamycin (Fisher Scientific), ampicillin (Fisher Scientific), stock solutions were filter sterilised[1] immediately.
3. Isopropyl β-D-1-thiogalactopyranoside (IPTG): 1 M made with distilled water and filter sterilised[1] immediately.

Method

A 37 °C overnight incubation of 3 mL of LB media with the appropriate antibiotic was inoculated with a single plate colony of cells. The following morning 30 µL was seeded into 3 mL (1:100 dilution) fresh LB media with antibiotic and grown at 37 °C to an OD_{600} 0.6-0.8. A 1 mL un-induced control sample was removed and the cells harvested by centrifugation at 15000 × g for 5 minutes. The remaining 2 mL of the cell culture was induced with 1 mM IPTG and re-incubated at 37 °C. After 3 hours, 1 mL of the culture was collected and centrifuged at 15000 × g for 5 minutes to harvest the induced cells. The cells were re-suspended in 50 µL 5 × SDS loading buffer and heated at 100 °C for 20 minutes. The samples were then centrifuged at 15000 × g for 5 minutes. The un-induced control was re-suspended in its supernatant while the supernatant of the induced samples were transferred into fresh bullet tubes and served as the soluble fractions. The induced sample pellets were re-suspended in 50 µL of fresh 5 × SDS loading buffer to serve as the insoluble fractions. All the samples were visualised on a SDS PAGE gel.

2.2.6 Western Blotting (Towbin *et al.* 1979; Burnette 1981)

Materials

1. Transfer buffer: 200 mM Glycine, 25 mM TRIS base, 10 % methanol, made up to the desired volume with distilled water and filter sterilised[1] immediately.
2. TBST: 50 mM TRIS·HCl pH 8.0, 150 mM NaCl, 0.1 % Tween, made up to the desired volume with distilled water and filter sterilised[1] immediately.

3. Bovine serum albumin (BSA) (Fisher Scientific): 10 mg mL^{-1} made with distilled water and stored at -20 °C.
4. Two-part developing solution: Part 1; 2.5 mM luminol, 0.5 mM p-coumaric acid, 100 mM TRIS·HCl pH 8.0, made up to the desire volume in distilled water. Part 2; 0.05 % v/v hydrogen peroxide, 100 mM TRIS·HCl pH 8.0, made up to the desire volume in distilled water.

Method

Samples were run on a SDS PAGE gel (2.1.2). Rather than being stained with coomassie blue staining solution, the gel was placed onto a nitrocellulose membrane which had been pre-soaked along with two pieces of blotting paper and two sponge sheets in transfer buffer. The nitrocellulose was placed between the two pieces of blotting paper which in turn were sandwiched between the two sponge sheets. The sandwich was then transferred into an electro-transfer box and run for 2 hours at 300 mA. Once transferred, the nitrocellulose membrane was washed with three 5 minutes washes with TBST at room temperature. It was then incubated with a 1° antibody in TBST with 5 % BSA at room temperature for 2 hours or 4 °C overnight. Next the membrane had three consecutive 5 minute washes with TBST at room temperature before incubation with a 2° horseradish peroxidise (HRP) conjugated antibody in TBST with 0.5 % BSA at room temperature for 1 hour. To finish, the membrane underwent a further three 5 minutes washes in TBST at room temperature. The blot was developed using a two-part developing solution and incubating for 1 minute before detecting with a Fujifilm LAS-3000 refrigerated camera at -20 °C under a 1 minute exposure time.

2.2.7 Growth and Cell Harvest

Materials

1. Lysogeny Broth (LB) media mix (Fisher Scientific): 1.5 % w/v tryptone, 0.5 % w/v yeast extract, 1 % w/v NaCl, made with distilled water and autoclave sterilised[1] immediately.

2. 2 × YT media mix (Fisher Scientific): 1.6 % w/v Tryptone, 1 % w/v yeast extract, 0.5 % w/v NaCl, made with distilled water and autoclave sterilised[1] immediately.
3. Antibiotic (Table 2.2): Kanamycin (Fisher Scientific), ampicillin (Fisher Scientific), stock solutions were filter sterilised[1] immediately.
4. IPTG: 1 M made with distilled water and filter sterilised[1] immediately.

Method

A 37 °C overnight incubation of LB media with the appropriate antibiotic was inoculated with a single plate colony of cells. The following morning the overnight seed culture was used in a 1:100 inoculation of the desired fresh media (LB or 2 × YT) containing the appropriate antibiotic. The culture was grown at 37 °C with agitation to an OD_{600} of 0.6-0.8 at which point the cells were induced with a suitable IPTG concentration. After a 3 hours expression period at 37 °C, or overnight at 20 °C, the cells were harvested by centrifugation at 4000 × g for 15 minutes at 4 °C and stored at -20 °C.

2.2.8 Auto-Induction Growth and Harvest (Hanahan *et al.* 1991)

Materials

1. ZY media: 1 % w/v tryptone, 0.5 % w/v yeast extract, made with distilled water and autoclave sterilised[1] immediately.
2. $MgSO_4$: 1M made with distilled water and filter sterilised[1] immediately.
3. 1000 × Metal mix: 50 mM $FeCl_3 \cdot 6H_2O$, 20 mM $CaCl_2$, 10 mM $MnCl_2 \cdot 4H_2O$, 10 mM $ZnSO_4 \cdot 7H_2O$, 2 mM $CoCl_2 \cdot 6H_2O$, 2 mM $CuCl_2 \cdot 2H_2O$, 2 mM $NiCl_2 \cdot 6H_2O$, 2 mM $Na_2MoO_4 \cdot 2H_2O$, 2 mM $Na_2SeO_3 \cdot 5H_2O$, 2 mM H_3BO_3, made with distilled water and filter sterilised[1] immediately.
4. 20 × NPS: 0.5M $(NH_4)_2SO_4$, 1M KH_2PO_4, 1M Na_2HPO_4, made with distilled water and autoclave sterilised[1] immediately.
5. 50 × 5052: 25 % w/v Glycerol, 2.5 % w/v glucose, 10 % w/v α-lactose, made with distilled water and autoclave sterilised[1] immediately.

Method

A 37 °C over-day incubation of 8 mL of ZYP-0.8G media (ZY media, 1 mM $MgSO_4$, 1 × Metal Mix, 0.8 % Glucose, 1 × NPS) with the appropriate antibiotic was inoculated with a single plate colony of cells. That evening, 4 mL of the starter culture was used to inoculate 400 mL ZYP-5052 (ZY media, 1 mM $MgSO_4$, 1 × metal mix, 1 × 5052, 1 × NPS) with the appropriate antibiotic and grown at 20 °C overnight. The following day the cells were harvested by centrifugation at 4000 × g for 15 minutes and stored at -20 °C.

2.2.9 Labelled Growth and Cell Harvest

Materials

1. Lysogeny Broth (LB) media mix (Fisher Scientific): 1.5 % w/v Tryptone, 0.5 % w/v yeast extract, 1 % w/v NaCl, made with distilled water and autoclave sterilised[1] immediately.
2. M9 minimal media: 0.6 % w/v Na_2HPO_4, 0.3 % w/v KH_2PO_4, 0.05 % w/v NaCl, made with distilled water and autoclave sterilised[1] immediately. Then add final reagents; 0.25 % w/v $(NH_4)_2SO_4$, 0.25 % w/v D(+) glucose, 1 mM $MgSO_4$, 1 mM thiamine.
3. 1000 × Metal mix: 50 mM $FeCl_3 \cdot 6H_2O$, 20 mM $CaCl_2$, 10 mM $MnCl_2 \cdot 4H_2O$, 10 mM $ZnSO_4 \cdot 7H_2O$, 2 mM $CoCl_2 \cdot 6H_2O$, 2 mM $CuCl_2 \cdot 2H_2O$, 2 mM $NiCl_2 \cdot 6H_2O$, 2 mM $Na_2MoO_4 \cdot 2H_2O$, 2 mM $Na_2SeO_3 \cdot 5H_2O$, 2 mM H_3BO_3, made with distilled water and filter sterilised[1] immediately.
4. Antibiotic (Table 2.2): Kanamycin (Fisher Scientific), ampicillin (Fisher Scientific), stock solutions were filter sterilised[1] immediately.
5. IPTG: 1 M made with distilled water and filter sterilised[1] immediately.

Method

A 37 °C over-day incubation of 5 mL of LB media with the appropriate antibiotic was inoculated with a single plate colony of cells. That evening 1 mL of the over-day culture was used to inoculate a 37 °C overnight incubation of 100 mL M9 minimal media containing 1 × metal mix, the appropriate antibiotic, and the desired isotope. The following morning the 100 mL overnight culture was added to 700 mL of fresh M9 minimal media

containing 1 × metal mix, the appropriate antibiotic, and the desired isotope. The culture was grown at 37 °C to an OD_{600} of 0.6-0.8 at which point the cells were induced with a suitable IPTG concentration. After a 3 hours expression period at 37 °C the cells were harvest by centrifugation at 4000 × g for 15 minutes at 4 °C and stored at -20 °C.

2.2.10 Sonication

Materials

1. Sonication buffer: 1 × Phosphate Buffered Saline (PBS) (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.4 mM K_2HPO_4), 1 mM DTT, 1 mM benzamidine, 0.1 mM phenylmethanesulphonylfluoride (PMSF), made up to the desired volume in distilled water.

Method

The cell pellet was thawed at room temperature, re-suspended in 35 mL sonication buffer and sonicated with a Sonics Vibra-Cell VCX-500 sonicator for 3.3s on and 9.9s off, at 40 % amplitude for 20 minutes at 4 °C. The cell lysate was then centrifuged at 50000 × g for 1 hour at 4 °C to pellet the insoluble fraction. The soluble fraction was decanted and kept at 4 °C prior to purification.

2.2.11 Resolubilisation with Sarkosyl

Materials

1. Sonication buffer: 1 × PBS (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.4 mM K_2HPO_4), 1 mM DTT, 1 mM benzamidine, 0.1 mM phenylmethanesulphonylfluoride (PMSF), made up to the desired volume in distilled water.
2. 10 % Sarkosyl: Made with distilled water.
3. Triton X-100.

4. 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS).

Method

The cell pellet was thawed at room temperature, re-suspended in 35 mL sonication buffer and sonicated for 3.3s on and 9.9s off, at 40 % amplitude for 30 minutes at 4 °C. The cell lysate was then centrifuged at 50000 × g for 1 hour at 4 °C to pellet the insoluble fraction. The soluble fraction was removed and the insoluble pellet was resuspended in 5 mL 10 % sarkosyl and re-sonicated. The sonicate was centrifuged at 50000 × g for 30 minutes. The supernatant was decanted off and diluted to 2 % Sarkosyl, made to 4 % Triton X-100 and 30 mM CHAPS. The sample was kept at 4 °C prior to purification.

2.2.12 Glutathione Affinity Chromatography

Materials

1. 5 mL GSTrap FF column (GE Healthcare).
2. Distilled water, filter sterilised[1] and vacuum degassed prior to use.
3. TRIS buffer A: 50 mM TRIS·HCl pH 8.0, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
4. TRIS buffer B: 50 mM TRIS·HCl pH 8.0, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10 mM glutathione, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
5. 20 % Ethanol: Made with distilled water, filter sterilised[1] and vacuum degassed prior to use.
6. PreScission protease (GE Healthcare).

Method

The sample was loaded into the superloop and purified as follows. At 4 °C and at a rate of 5 mL min^{-1} the column was pre-equilibrated with three column volumes of distilled water then three column volumes of TRIS A buffer. Collecting 4 mL fractions the sample was injected in TRIS buffer A followed by three column volumes of TRIS buffer A to wash

through all non-bound protein. An on column cleavage was performed by pausing the programme manually injecting 400 U of PreScission protease and incubating at 4 °C overnight. The following morning the cleaved protein was washed off the column with three column volumes of TRIS buffer A. The bound glutathione S-transferase (GST) moiety was eluted with three column volumes TRIS buffer B. Without collecting fractions, the column was washed with three column volumes of distilled water followed by three column volumes of 20 % ethanol for storage purposes. The fractionated sample was run on a SDS PAGE gel (2.1.2) and any protein containing fractions were pooled and stored at 4 °C.

2.2.13 Nickel Affinity Chromatography (Porath *et al.* 1975; Sulkowski 1989)

Materials

1. Ni-NTA resin (QIAGEN).
2. Distilled water, filter sterilised[1] and vacuum degassed prior to use.
3. Disodium phosphate buffer: 50 mM Na_2HPO_4 pH 7.5, 500 mM NaCl, 1 mM DTT, 10 % (v/v) Glycerol, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
4. HEPES buffer: 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulphonic acid (HEPES) pH 7.5, 500 mM NaCl, 1 mM DTT, 10 % (v/v) Glycerol, 200 mM imidazole, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
5. 20 % Ethanol: Made with distilled water, filter sterilised[1] and vacuum degassed prior to use.

Method

The sample was loaded onto the superloop and purified as follows. At 4 °C and at a rate of 5 mL min^{-1} the column was pre-equilibrated with three column volumes of distilled water then three column volumes of disodium phosphate buffer . Collecting 4 mL fractions the sample was injected in disodium phosphate buffer followed by three column volumes of

disodium phosphate buffer to wash through all non-bound protein. The bound protein was eluted with three column volumes HEPES buffer. Without collecting fractions, the column was washed with three column volumes of distilled water followed by three column volumes of 20 % ethanol for storage purposes. The fractionated sample was run on a SDS PAGE gel (2.1.2); any protein containing fractions were pooled and stored at 4 °C.

2.2.14 Diethylaminoethyl (DEAE) Cellulose Chromatography

Materials

1. 5 mL HiTrap DEAE FF column (GE Healthcare).
2. Distilled water, filter sterilised[1] and vacuum degassed prior to use.
3. TRIS buffer A: 20 mM TRIS·HCl pH 8.0, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
4. TRIS buffer B: 20 mM TRIS·HCl pH 8.0, 250 mM NaCl, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
5. 20 % Ethanol: Made with distilled water, filter sterilised[1] and vacuum degassed prior to use.

Method

The sample was loaded into the superloop and purified as follows. At 4 °C and at a rate of 5 mL min^{-1} the column was pre-equilibrated with three column volumes of distilled water then three column volumes of TRIS buffer A. Collecting 4 mL fractions the sample was injected in TRIS buffer A followed by three column volumes of TRIS buffer A to wash through all non-bound protein. Protein was eluted during a 0-100 % gradient of TRIS buffer B against TRIS buffer A. Without collecting fractions, the column was washed with three column volumes of distilled water followed by three column volumes of 20 % ethanol for storage purposes. The fractionated sample was run on a SDS PAGE gel (2.1.2) and any protein containing fractions were pooled and stored at 4 °C.

2.2.15 Desalting Chromatography

Materials

1. 5 mL HiTrap Desalting column (GE Healthcare).
2. Distilled water, filter sterilised[1] and vacuum degassed prior to use.
3. TRIS buffer: 50 mM TRIS·HCl pH 8.0, 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
4. 20 % Ethanol: Made with distilled water, filter sterilised[1] and vacuum degassed prior to use.

Method

The sample was loaded into the superloop and purified as follows. At 4 °C and at a rate of 5 mL min^{-1} the column was pre-equilibrated with three column volumes of distilled water then three column volumes of TRIS buffer A. Collecting 4 mL fractions the sample was injected washed through in TRIS buffer A. Without collecting fractions, the column was washed with three column volumes of distilled water followed by three column volumes of 20 % ethanol for storage purposes. The fractionated sample was run on a SDS PAGE gel (2.1.2); any protein containing fractions were pooled and stored at 4 °C.

2.2.16 Heparin Affinity Chromatography

Materials

1. 5 mL HiTrap Heparin HP column (GE Healthcare).
2. Distilled water, filter sterilised[1] and vacuum degassed prior to use.
3. HEPES buffer A: 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulphonic acid (HEPES) pH 7.5, 1 mM DTT, 10 % (v/v) Glycerol, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.

4. HEPES buffer B: 20 mM HEPES pH 7.5, 1 mM DTT, 500 mM NaCl, 10 % (v/v) Glycerol, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
5. 20 % Ethanol: Made with distilled water, filter sterilised[1] and vacuum degassed prior to use.

Method

The sample was loaded into the superloop and purified as follows. At 4 °C and at a rate of 5 mL min^{-1} the column was pre-equilibrated with three column volumes of distilled water then three column volumes of HEPES buffer A. Collecting 4 mL fractions the sample was injected in HEPES buffer A followed by three column volumes of HEPES buffer A to wash through all non-bound protein. Protein was eluted with HEPES buffer B. Without collecting fractions, the column was washed with three column volumes of distilled water followed by three column volumes of 20 % ethanol for storage purposes. The fractionated sample was run on a SDS PAGE gel (2.1.2); any protein containing fractions were pooled and stored at 4 °C.

2.2.17 Size Exclusion Chromatography

Materials

1. HiLoad 26/60 Superdex 200 Prep Grade column (GE Healthcare).
2. Distilled water, filter sterilised[1] and vacuum degassed prior to use.
3. HEPES buffer: 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulphonic acid (HEPES) pH 7.5, 1 mM DTT, 10 % (v/v) Glycerol 50 mM NaCl, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
4. 20 % Ethanol: Made with distilled water, filter sterilised[1] and vacuum degassed prior to use.

Method

The sample was loaded into the superloop and purified as follows. At 4 °C and at a rate of 3 mL min^{-1} the column was pre-equilibrated with one column volume of distilled water then one column volume of HEPES buffer. Collecting 4 mL fractions the sample was injected in size exclusion column running buffer followed by one and a half column volumes of HEPES buffer to resolve and elute the protein. Without collecting fractions, the column was washed with one column volume of distilled water followed by one and a half column volumes of 20 % ethanol for storage purposes. The fractionated sample was run on a SDS PAGE gel (2.1.2); any protein containing fractions were pooled and stored at 4 °C.

2.2.18 Ammonium Sulphate Fractionation

Materials

1. $(NH_4)_2SO_4$.
2. TRIS buffer: 50 mM TRIS·HCl pH 8.0, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.

Method

Protein samples were made to 10 % $(NH_4)_2SO_4$ and incubated with agitation at room temperature for 20 minutes. The sample was then centrifuged at 50000 × g for 15 minutes. The supernatant was decanted into a fresh centrifuge vessel and kept for the next step while the pellet was re-suspended in a suitable volume of TRIS buffer. This was repeated for the supernatant increasing the $(NH_4)_2SO_4$ concentration by 10 % each round to a maximum of 50 %. Samples of 0, 10, 20, 30, 40, and 50 % $(NH_4)_2SO_4$ as well as the last supernatant were run on a SDS PAGE gel (2.1.2).

2.2.19 Growth and Purification of GST-PP-HsdR

Materials

1. LB agar (Lennox L Agar) plates mix (Fisher Scientific): 1 % w/v Tryptone, 0.5 % w/v yeast extract, 5 % w/v NaCl, 1 % w/v agar, made with distilled water and autoclave sterilised[1] immediately. Once sufficiently cooled to touch, ampicillin was added to a final concentration of 100 µg mL^{-1} and the molten agar was poured into culture plates to a depth of approximately 5 mm. Once set the plates were stored at 4 °C until required for a maximum of two weeks.
2. Lysogeny Broth (LB) media mix (Fisher Scientific): 1.5 % w/v Tryptone, 0.5 % w/v yeast extract, 1 % w/v NaCl, made with distilled water and autoclave sterilised[1] immediately.
3. IPTG: 1 M made with distilled water and filter sterilised[1] immediately.
4. TRIS buffer A: 50 mM TRIS·HCl pH 8.0, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
5. TRIS buffer B: 50 mM TRIS·HCl pH 8.0, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10 mM glutathione, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
6. PreScission protease (GE Healthcare).
7. 100 % Glycerol.

Method

The purification of HsdR was carried out following the unpublished work performed by Dr James Youell (University of Portsmouth, Hampshire, UK).

The recombinant pGEX-6P-1 plasmid harbouring the *HsdR* or *PreScission Protease* (*PP*) gene was transformed (2.2.3) into BL21 (DE3) [pLysS] *E. coli* cells (Stratagene) (Table 2.2), and grown overnight on LB agar plates supplemented with 100 µg mL^{-1} ampicillin at 37 °C.

37 °C overnight seed cultures were set up, containing single colonies of BL21 (DE3) [pLysS] transformants in LB media with 100 µg mL^{-1} ampicillin. These were used a 1:100 inoculation of LB media growth containing 100 µg mL^{-1} ampicillin and grown at 37 °C to an OD_{600} of 0.6-0.8 at which point the cells were induced with 1 mM IPTG. After an overnight expression period at 24 °C the cells were harvested by centrifugation at 4000 × g for 15 minutes at 4 °C and stored at -20 °C.

The cell pellet was sonicated as described in 2.2.10 and the supernatant was additionally ultra-centrifuged at 200000 × g for 30 minutes. The supernatant was loaded into the superloop and purified as follows. At 4 °C and at a rate of 5 mL min^{-1} the column was pre-equilibrated with three column volumes of distilled water then three column volumes of TRIS A buffer. Collecting 4 mL fractions the sample was injected in TRIS buffer A followed by three column volumes of TRIS buffer A to wash through all non-bound protein. An on column cleavage of HsdR-PP-GST was performed by pausing the programme, and manually injecting 400 U of PreScission protease and incubating at 4 °C overnight. The following morning the cleaved protein was washed off the column with three column volumes of TRIS buffer A. The bound glutathione S-transferase (GST) moiety was eluted with three column volumes TRIS buffer B. Without collecting fractions, the column was washed with three column volumes of distilled water followed by three column volumes of 20 % ethanol for storage purposes. Fraction samples were run on a SDS PAGE gel (2.1.2) and any protein containing fractions were pooled, concentrated with a centrifugal concentrator, and made to 50 % glycerol for long term storage at -20 °C.

2.2.20 Purification of MTase

Materials

1. LB agar (Lennox L Agar) plates mix (Fisher Scientific): 1 % w/v Tryptone, 0.5 % w/v yeast extract, 5 % w/v NaCl, 1 % w/v agar, made with distilled water and autoclave sterilised[1] immediately. Once sufficiently cooled to touch, ampicillin was added to a final concentration of 100 µg mL^{-1} and the molten agar was poured into culture plates to a depth of approximately 5 mm. Once set the plates were stored at 4 °C until required for a maximum of two weeks.
2. Lysogeny Broth (LB) media mix (Fisher Scientific): 1.5 % w/v Tryptone, 0.5 % w/v yeast extract, 1 % w/v NaCl, made with distilled water and autoclave sterilised[1] immediately.
3. IPTG: 1 M made with distilled water and filter sterilised[1] immediately.
4. TRIS buffer A: 50 mM TRIS·HCl pH 8.0, 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.

5. TRIS buffer B: 50 mM TRIS·HCl pH 8.0, 1 M NaCl, 0.1 mM EDTA, 1 mM DTT, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
6. TRIS buffer C: 50 mM TRIS·HCl pH 8.0, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, made up to the desired volume with distilled water, filter sterilised[1] and vacuum degassed prior to use.
7. 100 % Glycerol.

Method

Purification of MTase was performed as described by Taylor (Taylor *et al.* 1992) with minor modification based on personal correspondence with Dr Peter Coxhead (University of Portsmouth, Hampshire, UK). In short, the pJS4M (Patel *et al.* 1992) plasmid encoding wild-type MTase was transformed (2.2.3) into JM109 (DE3) cells (Promega) (Table 2.2) and grown overnight on LB agar plates supplemented with 100 µg mL^{-1} ampicillin at 37 °C.

37 °C overnight seed cultures were set up, containing single colonies of JM109 (DE3) transformants in LB media with 100 µg mL^{-1} ampicillin. These were used a 1:100 inoculation of 2 × YT media growth containing 100 µg mL^{-1} ampicillin and grown at 37 °C to an OD_{600} of 0.6-0.8 at which point the cells were induced with 1 mM IPTG. After an 6 hour expression period at 37 °C the cells were harvested by centrifugation at 4000 × g for 15 minutes at 4 °C and stored at -20 °C.

The cell pellet was sonicated as described in 2.2.10 and the supernatant was additionally ultra-centrifuged at 200000 × g for 30 minutes. The supernatant was then made to 500 mM NaCl and 10 mg mL^{-1} of protamine sulphate, gently mixed for 30 minutes at 4 °C and centrifuged at 27000 × g for 30 minutes. The supernatant was then made to 70 % ammonium sulphate, gently mixed for 30 minutes at 4 °C and centrifuged at 12000 × g for 30 minutes. The supernatant was discarded and the pellet was resuspended in TRIS Buffer A, and then desalted using a HiTrap Desalting column (GE Healthcare). The desalted fractions were then loaded on HiTrap DEAE columns (GE Healthcare) and eluted over a 100 mL gradient of 50 mM to 500 mM NaCl using TRIS Buffer B. Fractions containing MTase were pooled and desalted again, this time into TRIS Buffer C. The desalted fractions were pooled and loaded onto a HiTrap Heparin column (GE Healthcare) and

eluted over a 100 mL gradient of 100 mM to 1 M NaCl using TRIS Buffer B. Fraction samples were run on a SDS PAGE gel (2.1.2) and any protein containing fractions were pooled, concentrated to with a centrifugal concentrator, and made to 50 % glycerol for long term storage at -20 °C.

2.2.21 UV-Visible Spectrophotometry

Protein and DNA concentrations were measured using a Nanodrop 2000 spectrophotometer (Thermo Scientific) at 280 nm and 260 nm wavelengths respectively. Concentrations were calculated via the Beer-Lambert law:

$$A = \varepsilon cl$$

Where A is the absorbance at a given wavelength, ε is the molar extinction coefficient measured in per molar concentration per centimetre (M^{-1} cm^{-1}), c is the molar concentration of the measured sample and l is the sample path length.

Duplex DNA concentrations were estimated based on a 50 µg mL^{-1} sample having an absorbance of 1 absorbance unit (AU) at 260 nm (Sambrook *et al.* 2001). Protein concentrations were either estimated based on a 1 mg mL^{-1} sample having an absorbance of 1 AU, or calculated more accurately using their specific molar extinction coefficients and molecular weights.

2.3 Biochemical Characterisation Protocols

2.3.1 Substrate Annealing (Özsoy *et al.* 2003)

Materials

1. Oligonucleotides (Invitrogen) (Appendix 1): 100 µM made with distilled water.

2. Annealing buffer: 7 mM TRIS·HCl pH 8.0, 6.6 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.

Method

Substrates annealing follows the protocol described by Özsoy *et al.* (2003) with minor modification. Equimolar ratios of the substrates component oligonucleotides were incubated together in annealing buffer at 95 °C for 20 minutes then allowed to slowly cool to room temperature and stored at -20 °C.

2.3.2 Triplex Formation (Firman *et al.* 2000)

Materials

1. Triplex forming oligonucleotide (TFO) (Invitrogen):

   TTC TTT TCT TTC TTC TTT CTT T

   100 µM made with distilled water.

2. DNA duplex.
3. Triplex buffer: 25 mM 2-(N-morpholino)ethanesulphoic acid (MES) pH 5.5, 10 mM $MgCl_2$, made with distilled water and autoclave sterilised[1] immediately.

Method

The duplex substrate, prepared as described in 2.3.1, was incubated with the TFO at an equimolar ratio in Triplex buffer at 57 °C for 20 minutes then allowed to slowly cool to room temperature and stored at -20 °C.

2.3.3 Helicase Assay (Özsoy *et al.* 2003)

Materials

1. Substrates (Table 2.3).
2. Helicase.
3. 10 × Assay buffer: 250 mM TRIS·HCl pH 8.0, 200 mM NaCl, 50 mM $MgCl_2$, and 10 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.
4. Adenosine triphosphate (ATP) (Fisher Scientific): 20 mM made with distilled water and stored at -80 °C.
5. Bovine serum albumin (BSA) (Fisher Scientific): 10 mg mL^{-1} made with distilled water and stored at -20 °C.
6. 5 × Termination buffer: 1.5 % SDS, 50 mM EDTA, 25 % (v/v) Glycerol, made up to the desired volume with distilled water and filter sterilised[1] immediately.

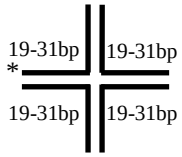
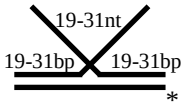
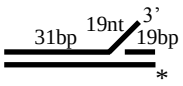
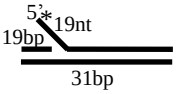
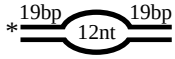
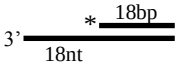
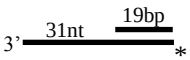
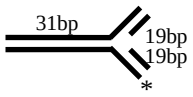
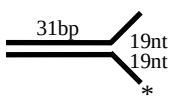
Method

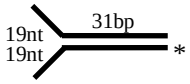
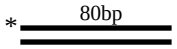
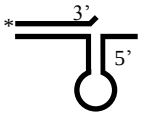
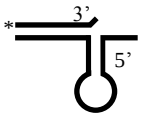
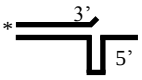
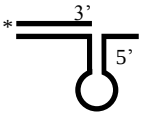
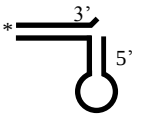
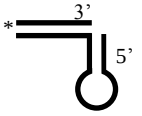
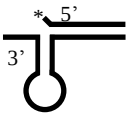
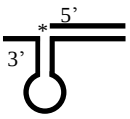
Reactions set up in bullet tubes containing a 0.1 µM phosphoflorescent substrate, the necessary concentration of helicase, 1 mM ATP, and 0.8 mg mL^{-1} BSA, in 1 × assay buffer were incubated at 37 °C for 1 hour. The reactions was stopped by the addition of 5 × termination buffer and incubated at 37 °C for 5 minutes. The reactions were run on a SDS PAGE gel (2.1.2) and visualised using a Fujifilm FLA-5000 phosphoimager with a blue laser (473 nm) at 550 V and a green filter (532 nm).

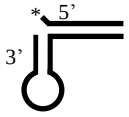
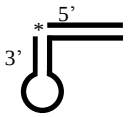
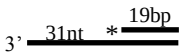
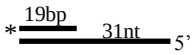
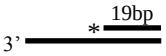
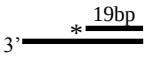
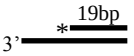
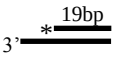
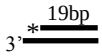
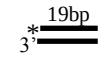
Gel densitometry was performed using the image processing and analysis software ImageJ (Schneider *et al.* 2012). The following equation was used to quantify the degree of unwinding and used for graphical representation generated using GraFit5 (Erithacus Software):

$$Primary\ Substrate\ = \frac{(P - P_B\ )}{(S - S_B) + \ (P - P_B)} \times\ 100$$

Where P represents the unwound product and P_B represents any product background (unannealed product present in the 0min time-point control sample), S represents the primary substrate and S_B represents any thermal degradation of the primary substrate (calculated from the difference in primary substrate between the 0min and final time-point control samples).

Number	Name	Description	Structure	Oligonucleotides Used
1	Holliday Junction	Holliday junction with 4 blunt ends		1*+2+3+4
2	3 Strand Junction	3 strand junction with 2 blunt ends and 2 single stranded ends		1+3+4*
3	3’ Flap	Blunt ended duplex with an incorporated 3’ flap		1+4*+7
4	5’ Flap	Blunt ended duplex with an incorporated 5’ flap		1+4*+8
5	4 bp Bubble	Blunt ended duplex with a 4 nucleotide bubble		1*+5
6	12 bp Bubble	Blunt ended duplex with a 12 nucleotide bubble		1*+6
7	36/18 Partial Duplex	Partial duplex with 1 blunt end and an 18 nucleotide single stranded overhang		9*+10
8	50/19 Partial Duplex	Partial duplex with 1 blunt end and a 31 nucleotide single stranded overhang		4*+7
9	3 Way Junction	3 way junction with 3 blunt duplex ends		1+4*+7+8
10	Fork 1	Duplex with a blunt end and a 3’ forked end		1+4*

11	Fork 2	Duplex with a blunt end and a 5’ forked end		3+4*
12	80 bp Blunt Duplex	80 bp dsDNA with 2 blunt ends		11*+12
13	Hairpin Duplex (HPD) 1	5’ hairpin with a 5’ single-stranded overhang and a 1 nucleotide 3’ flap		13+14*
14	HPD2	5’ hairpin with a 5’ single-stranded overhang and a 1 nucleotide 3’ flap, alternative hairpin sequence		14*+15
15	HPD3	5’ hairpin with a 5’ single-stranded overhang and a 1 nucleotide 3’ flap, deleted loop sequence		14*+16
16	HPD4	5’ hairpin with a 5’ single-stranded overhang		13+17*
17	HPD5	5’ hairpin with a 1 nucleotide 3’ flap		14*+18
18	HPD6	5’ hairpin no single-stranded overhang		17*+18
19	HPD7	3’ hairpin with a 3’ single-stranded overhang and a 1 nucleotide 5’ flap		19+20*
20	HPD8	3’ hairpin with a 3’ single-stranded overhang		19+21*

21	HPD9	3’ hairpin with a 1 nucleotide 5’ flap		20*+22
22	HPD10	3’ hairpin no single-stranded overhang		21*+22
23	50/19 Partial Duplex (3’-5’)	Partial duplex with 1 blunt end and a 3’ 31 nucleotide single stranded overhang		4+7*
24	50/19 Partial Duplex (5’-3’)	Partial duplex with 1 blunt end and a 5’ 31 nucleotide single stranded overhang		23*+24
25	45/19 Partial Duplex (3’-5’)	Partial duplex with 1 blunt end and a 3’ 26 nucleotide single stranded overhang		4*+25
26	40/19 Partial Duplex (3’-5’)	Partial duplex with 1 blunt end and a 3’ 21 nucleotide single stranded overhang		4*+26
27	35/19 Partial Duplex (3’-5’)	Partial duplex with 1 blunt end and a 3’ 16 nucleotide single stranded overhang		4*+27
28	30/19 Partial Duplex (3’-5’)	Partial duplex with 1 blunt end and a 3’ 11 nucleotide single stranded overhang		4*+28
29	25/19 Partial Duplex (3’-5’)	Partial duplex with 1 blunt end and a 3’ 6 nucleotide single stranded overhang		4*+29
30	20/19 Partial Duplex (3’-5’)	Partial duplex with 1 blunt end and a 3’ 1 nucleotide single stranded overhang		4*+30

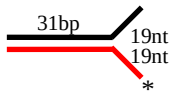
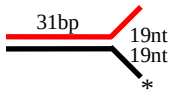
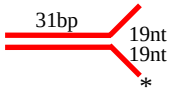
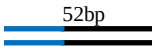
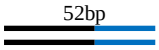
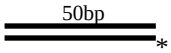
31	DNA:RNA Hybrid Fork	DNA:RNA hybrid duplex with a blunt end and a 3' forked end		1+31*
32	RNA:DNA Hybrid Fork	RNA:DNA hybrid duplex with a blunt end and a 3' forked end		4*+32
33	dsRNA Hybrid Fork	dsRNA duplex with a blunt end and a 3' forked end		31*+32
34	5' TFO duplex	52 bp duplex with a 5' triplex binding site		33+34
35	3' TFO duplex	52 bp duplex with a 3' triplex binding site		35+36
36	ds50mer Fork 1	Full complimentary sequence of Fork 1		4*+37

Table 2.3 **Helicase Substrates.** The substrates are constructed from the oligonucleotides detailed in Appendix 1. The left most column contains the substrates assigned number, while the second and third columns contain their name and description respectively. The fourth column contains a schematic of each of the substrates, and the fifth column has the constituent oligonucleotide numbers (Appendix 1). The structures are depicted with the upper-left most strand in the 5'-3' (left to right) orientation. Oligonucleotides marked with a * indicate that they are 5' flouroscein labelled. The red lines in the schematics of substrates 31 to 33 depicted the RNA strands, while the blue portions of substrates 34 and 35 highlight the triplex binding sites.

2.3.4 BOXTO Helicase Assay

Materials

1. BOXTO-substrate adducts: prepared by incubating BOXTO with the substrate at a 5:1 molar ratio.

2. Helicase.
3. 10 × Assay buffer: 250 mM TRIS·HCl pH 8.0, 200 mM NaCl, 50 mM $MgCl_2$, and 10 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.
4. 10 × Negative control buffer: 250 mM TRIS·HCl pH 8.0, 200 mM NaCl, 50 mM EDTA, and 10 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.
5. Adenosine triphosphate (ATP) (Fisher Scientific): 20 mM made with distilled water and stored at -80 °C.
6. Bovine serum albumin (BSA) (Fisher Scientific): 10 mg mL^{-1} made with distilled water and stored at -20 °C.

Method

Three 10 µL reactions containing 2 × the necessary concentration of helicase and 1.6 mg mL^{-1} BSA, two in 2 × assay buffer and the third in 2 × negative control buffer were aliquot into a 96-well plate. 10 µL of the start solution containing 0.2 µM substrate and 2 mM ATP were added two one of the assay buffer samples and the negative control buffer sample. A third start solution with a boiled substrate, incubated at 95 °C for 20 minutes then rapidly cooled on ice to prevent reannealing, was added to the second assay buffer sample to serve as a positive control. The reaction was followed by exciting at 485/20 and detecting at 530/25 nm using a Bio-Tek FL600 fluorescence plate reader at 37 °C for 1 hour.

2.3.5 Phosphorylation of Helicase

Materials

1. Helicase.
2. Protein Kinase C (PKC) (Thermo Scientific).
3. PKC buffer: 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulphonic acid (HEPES) pH 7.5, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 1 mM ATP, made up to the desired volume with distilled water and filter sterilised[1] immediately.

4. Adenosine triphosphate (ATP) (Fisher Scientific): 20 mM made with distilled water and stored at -80 °C.

Method

Helicase was phosphorylated by incubating with 1ng of PKC for every 200 ng of helicase at room temperature for 30 minutes in PKC buffer with 1 mM ATP. The phosphorylated protein was then used in the helicase and ATPase assays.

2.3.6 Electrophoretic Mobility Shift Assay (EMSA) (Fried *et al.* 1981; Garner *et al.* 1981)

1. Substrates (Table 2.3).
2. Helicase.
3. 10 × Assay buffer: 250 mM TRIS·HCl pH 8.0, 200 mM NaCl, 50 mM $MgCl_2$, and 10 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.
4. Adenosine 5'-(β,γ-imido)triphosphate (AMP-PnP) (Fisher Scientific): 20 mM made with distilled water and stored at -80 °C.
5. Bovine serum albumin (BSA) (Fisher Scientific): 10 mg mL^{-1} made with distilled water and stored at -20 °C.

Method

Reactions set up in bullet tubes containing a 0.1 µM phosphoflorescent substrate, the necessary concentration of helicase, 1 mM AMP-PnP, and 0.8 mg mL^{-1} BSA, in 1 × assay buffer were incubated at room temperature for 5 minutes. The reactions were run on a native PAGE gel (2.1.3) and visualised using a Fujifilm FLA-5000 phosphoimager with a blue laser (473 nm) at 550 V and a green filter (532 nm).

Gel densitometry was performed using the image processing and analysis software ImageJ (Schneider *et al.* 2012). The following equation was used to fit the data to a single site saturation curve, where the amount of ligand bound is plotted as a function of the amount of free ligand, using GraFit5 (Erithacus Software):

$$[Bound] = \frac{Capacity \times [Free]}{K_d + [Free]}$$

Where [*Bound*] represents the concentration of the bound ligand, *capacity* is the interaction product maxima, [*Free*] is the concentration of the unbound ligand, and K_d is the dissociation constant.

2.3.7 Malachite Green ATPase Assay (Chan *et al.* 1986; D'Angelo *et al.* 2001)

Materials

1. Substrates (Table 2.3).
2. Helicase.
3. 10 × Assay buffer: 250 mM TRIS·HCl pH 8.0, 200 mM NaCl, 50 mM $MgCl_2$, and 10 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.
4. Adenosine triphosphate (ATP) (Fisher Scientific): 100 mM made with distilled water and stored at -80 °C.
5. Bovine serum albumin (BSA) (Fisher Scientific): 10 mg mL^{-1} made with distilled water and stored at -20 °C.
6. Ammonium molybdate tetrahydrate: 5.72 % w/v of 6 M HCl
7. Polyvinyl alcohol (PVA): 2.32 % w/v of boiling distilled water.
8. Malachite green: 0.0812 % w/v of distilled water.
9. Distilled water.

Method

10 µL time-points of a reverse time-course reaction, containing 2 × the necessary concentration of helicase and 1.6 mg mL^{-1} BSA in assay buffer were aliquot into a 96-well plate and incubated at 37 °C. At each reverse time-point, 10 µL of the start solution containing 2 µM substrate and 20 mM ATP in assay buffer was added. The reactions were stopped with 150 µL of a 1:1:2:2 mixture of ammonium molybdate, PVA, malachite green and distilled water at the 0 minute time-point. After a 10 minutes incubation at room-

temperature the inorganic phosphate (P_i) was detected using an Apollo-8 LB-912 spectrophotometer (Berthold Technologies) reading the absorbance at 590-650nm. The concentration of the inorganic phosphate (P_i) was determined by comparison against a standard line derived from serial dilutions of KH_2PO_4 (Figure 2.3).

2.3.8 Luciferase ATPase Assay

Materials

1. Substrates (Table 2.3).
2. Helicase.
3. 10 × Assay buffer: 250 mM TRIS·HCl pH 8.0, 200 mM NaCl, 50 mM $MgCl_2$, and 10 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.
4. Adenosine triphosphate (ATP) (Fisher Scientific): 20 mM made with distilled water and stored at -80 °C.
5. Bovine serum albumin (BSA) (Fisher Scientific): 10 mg mL^{-1} made with distilled water and stored at -20 °C.
6. Standard Reaction Solution: Prepared as described in the ATP Determination Kit (Invitrogen); 1.25 µg mL^{-1} luciferase with 0.5 mM D-luciferin in 25 mM Tricine (pH 7.8), 5 mM $MgSO_4$, 0.1 mM EDTA, 1 mM DTT, and 0.1 mM sodium azide, stored short term at 4 °C protected from light.
7. Distilled water.

Method

10 µL time-points of a reverse time-course reaction, containing 2 × the necessary concentration of helicase and 1.6 mg mL^{-1} BSA in assay buffer were aliquot into a 96-well plate and incubated at 37 °C. At each reverse time-point, 10 µL of the start solution containing 2 µM substrate and 2 mM ATP in assay buffer was added. The reactions were stopped with 100 µL Standard Reaction Solution at the 0 minute time-point and the luminescence was taken using a 1450 MicroBeta TriLux luminescence counter. The

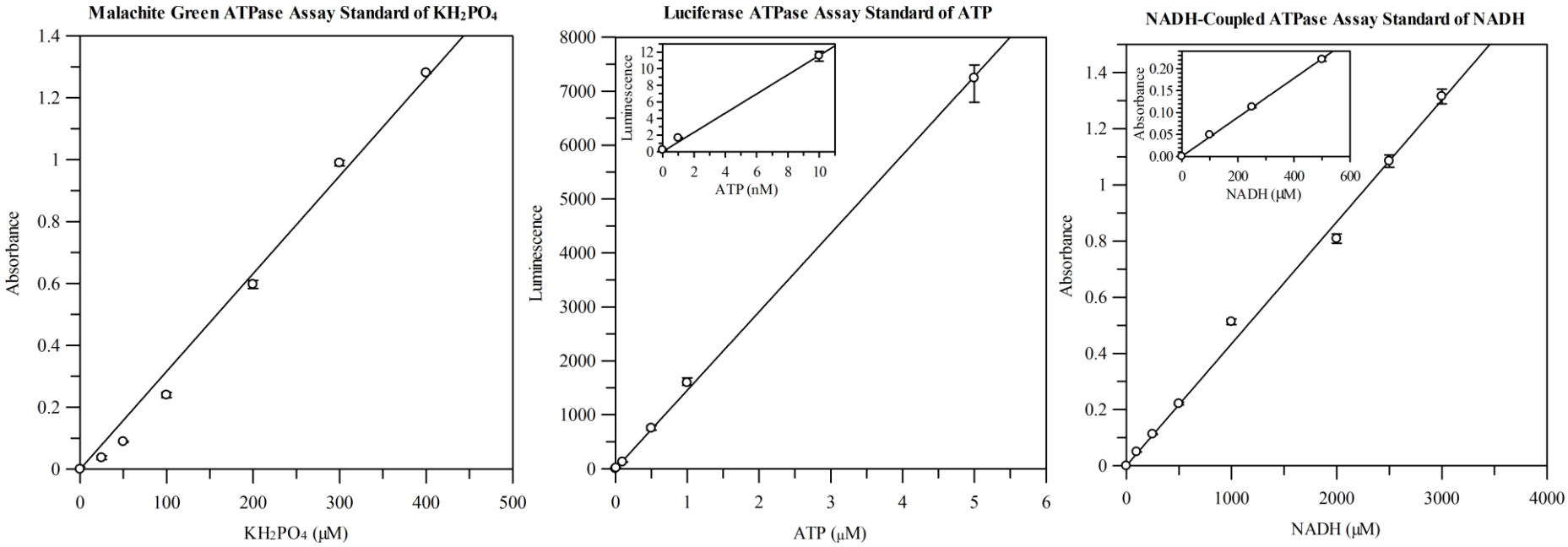


Figure 2.3 **ATPase Assay Standards.** Created with GraFit5 (Erithacus Software), the graphs show example standard lines for the malachite green, luciferase, and NADH-coupled ATPase assays. Derived from the serial dilutions of potassium dihydrogen phosphate (KH_2PO_4), the leftmost graph line equation can be used to fit the raw data obtained from the malachite green ATPase assay to elucidate the amount of inorganic phosphate released from the hydrolysis of ATP. The central graph shows an ATP standard line equation that can be used to calculate the amount of unused ATP from the raw data obtained from the luciferase ATPase assay. The rightmost graph was generated using NADH standards and can be used for a direct comparison for the amount of ATP hydrolysed during the NADH-coupled ATPase assay.

concentration of the ATP was determined by comparison against a standard line derived from serial ATP dilutions (Figure 2.3).

2.3.9 NADH-Coupled ATPase Assay (Kiianitsa *et al.* 2003)

Materials

1. Substrates (Table 2.3).
2. Helicase.
3. 10 × Assay buffer: 250 mM TRIS·HCl pH 8.0, 200 mM NaCl, 50 mM $MgCl_2$, and 10 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.
4. Adenosine triphosphate (ATP) (Fisher Scientific): 100 mM made with distilled water and stored at -80 °C.
5. Bovine serum albumin (BSA) (Fisher Scientific): 10 mg mL^{-1} made with distilled water and stored at -20 °C.
6. Phosphoenolpyruvate (PEP) (Sigma-Aldrich): 150 mM made with distilled water and stored at -20 °C.
7. 1000 U mL^{-1} Pyruvate kinase (PK)/L-lactate dehydrogenase (LDH) (Sigma-Aldrich).
8. Reduced β-Nicotinamide adenine dinucleotide (NADH) (Sigma-Aldrich): 30 mM made with distilled water and stored at -20 °C.
9. Distilled water.

Method

Reaction containing a 0.1 μM phosphoflorescent substrate, 5 mM ATP, 0.8 mg mL^{-1} BSA, 3 mM PEP, 20 U mL^{-1} PK/LDH, 2.5 mM NADH, and the appropriate volume of helicase in 1 × assay buffer were set up in bullet tubes. A 0 minute time-point was taken immediately by monitoring the absorbance at 339 nm at NADHs extinction coefficient of 6220 M^{-1} cm^{-1}, and at appropriate time-points thereafter. The concentration of the ATP hydrolysed was directly comparable to the decrease in NADH concentration, and was determined from a standard line derived from serial dilutions of NADH (Figure 2.3).

2.3.10 EcoR124I Cleavage Assay

Materials

1. pCFD30.
2. MTase.
3. HsdR.
4. Adenosine triphosphate (ATP) (Fisher Scientific): 20 mM made with distilled water and stored at -80 °C.
5. 10 × Cleavage Buffer: 500 mM Tris·HCl pH 8.0, 500 mM NaCl, 100 mM $MgCl_2$ and 10 mM DTT, made up to the desired volume with distilled water and filter sterilised[1] immediately.

Method

10 µL reactions containing a 10:1 HsdR to Mtase ratio to favour the R2 complex (2.5 µM HsdR and 250 nM Mtase), 2 mM ATP, and 2.5 nM pCFD30, in 1 × cleavage buffer were set up and incubated at 37 °C. The reactions were terminated after the desired time period by heat inactivation of the enzyme at 65 °C for 20 minutes. The reactions were run on a 0.8 % agarose gel to observe cleavage.

2.4 Structural Characterisation Protocols

2.4.1 Analytical Ultracentrifugation (AUC)

AUC experiments provide information on shape and molecular mass as well as stoichiometry and macromolecular interactions (Schuck 2005; Brown *et al.* 2006). Macromolecules will be ubiquitously dispersed throughout a given solution, however when centrifugal force was applied to the macromolecules their intrinsic tendency to settle out of their entrained solvent, called sedimentation, was accelerated. Sedimentation was opposed

by the frictional force acting upon the macromolecules, these opposing forces can be characterised in an AUC experiment through the Svedberg equation:

$$s = \frac{v}{\omega^2 r} = \frac{M(1 - \bar{v}\rho)}{N_A f}$$

Where s is the sedimentation coefficient, v is the rate of sedimentation, ω is the angular velocity, r is the radius (distance from the centre of rotation), M is the molar mass, $\bar{v}$ is the partial specific volume, ρ is the solvent density, N_A is Avogadro's constant, and f is the frictional coefficient. The sedimentation coefficient is expressed in Svedberg units (S), with 1 S being equal to exactly 10^{-13} seconds.

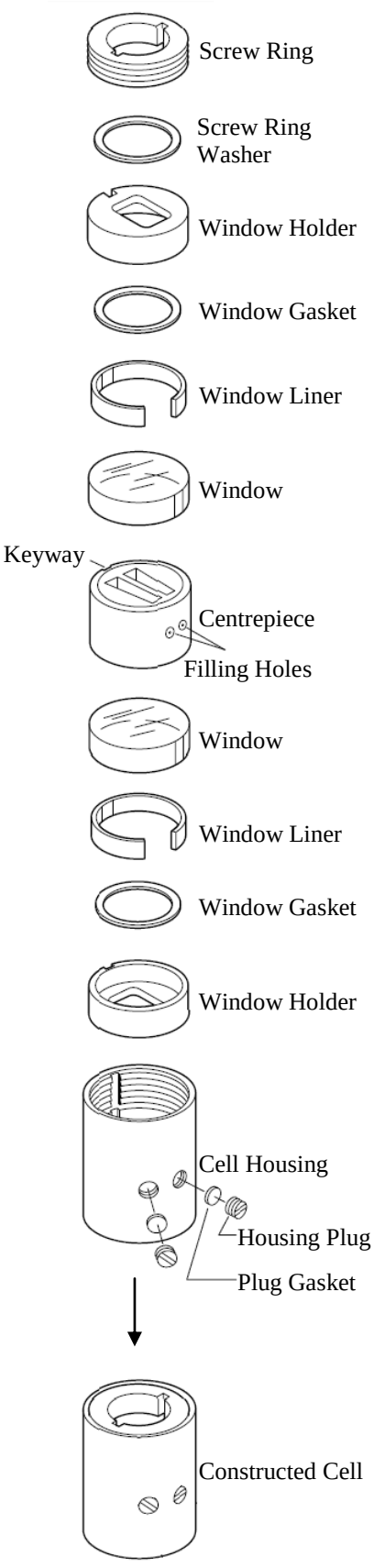
During an AUC experiment, the sedimentation of the macromolecules was followed by optically monitoring a sample cell. When studying proteins, they have distinct absorbance maxima at 200 and 280 nm; conferred primarily by peptide bonds and aromatic residues respectively. However, an individual protein species maximum absorbance can be altered by primary, secondary, tertiary, and quaternary structure, but was typically between (230 and 280 nm). This can be exploited during AUC analysis to locate their radial position in the sample cell using the appropriate light wavelength. The absorbance was then recorded relative to a reference cell.

AUC analysis was performed on protein samples at concentrations that gives an absorbance at 280 nm of 1 AU, at approximate rotor speeds based on their molecular weight as described by Holladay (1980) using a Beckman Coulter Optima XL-A Analytical Ultracentrifuge with an An-50 Ti Analytical Rotor (Figure 2.4).

2.4.2 1D and 2D Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR Spectroscopy can be used to determine the physical and chemical characteristics of a molecule by exploiting the magnetic properties of their constituent atomic nuclei (Keeler 2011).

Protons and neutrons within an atom spin. In many atoms, such as ^{13}C, the nucleus possesses an net spin (I) which can be determined when considering that; an atom with an



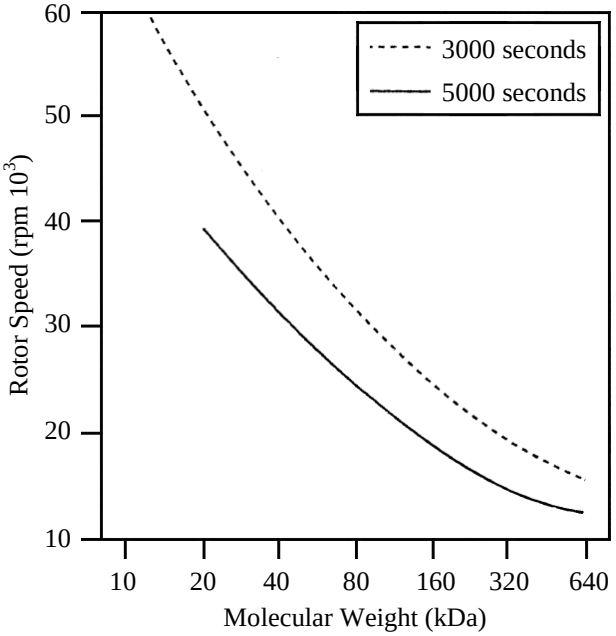


Figure 2.4 **Approximate Rotor speed and the Construction of an analytical ultracentrifugation cell.** Above, adapted from Holladay (1980), is a plot of suggested rotor speeds for two data collection periods times. Left, Adapted from the *An-50 Ti and An-60 Ti Analytical Rotor, Cells, and Counterbalance* manual from Beckman Coulter, the schematic shows the individual components which contribute to the sample cell. A window gasket, window liner, and quartz window are placed in that order into each of the window holders. One of these is then placed into the cell housing followed by the 12 mm Epon charcoal-filled centrepiece guided by its keyway. The second window holder is placed in above the centrepiece followed by the greased screw ring washer and the greased screw ring and tightened to 700 kPa. A 250 µL sample is injected into the left filling hole and a 200 µL solvent reference is injected into the right filling hole. The disparity between in the sample and reference volumes allows for discrimination between the two menisci. To finish, the plug gaskets and housing plugs are screwed into the cell housings.

even number of protons and neutrons has no net spin, an atom where the total number of protons and neutrons were odd has a half-integer net spin, and an atom which has an odd number of both protons and neutrons has a whole-integer net spin (Doucleff *et al.* 2011). A nucleus with a net spin has a fixed number of orientations based on:

$$2I + 1$$

Therefore a nucleus with an overall spin of ½ (^{1}H) has 2 possible orientations of equal energy. However, nuclei in each orientation can be separated into discrete energy levels by applying a magnetic field. Thermodynamics dictates that the nuclei population in the lower energy level is greater than that in the higher level. The extent of inequality is linked the degree of the separation with larger separation between the spin energy levels giving a greater population difference (Keeler 2011).

The nuclei in the lower energy population can be excited to the higher energy level, thus affecting their orientation, if a magnetic field which is greater than the energy difference (the transition energy) is applied. This is possible as the positive charged spinning nucleus generates its own small magnetic field and has an associated magnetic moment in the associated orientation. If the nucleus absorbs energy from an applied magnetic field the magnetic moment flips into the opposite orientation to oppose the field. Electrons around the nucleus shield it, so that the resonant frequency of the opposing nuclear magnetic field is not equal to that of the applied field.

The difference between the resonant and applied magnetic frequency is termed the chemical shift (Keeler 2011). The position and number of chemical shifts produced in an NMR experiment is affected by spin coupling, a small interaction between neighbouring groups of protons within a molecule. For example, a methylene (CH_2) group has two protons (^{1}H), each of which can has two possible orientations, either with or opposed to the applied magnetic field, giving four combinations; protons in both positions orientated with the applied field, both protons opposed, the proton in one position with and the other opposed, or the proton in one position opposed and the other with. Neither of the latter combinations has an effect on neighbouring proton groups. However, where both proton orientations were with the applied field, their own magnetic fields increase the field exerted on the neighbouring proton group, likewise if they oppose they reduce it. Therefore, the neighbouring proton group affectively receives three different magnetic

field frequencies, the uninterrupted applied frequency, and the increased and reduced frequencies. This results in a spectrum for the neighbouring proton group with a three peak profile, with a larger central peak corresponding to the two non-affecting proton combinations, and two half sized peaks up and downfield of the central peak, correlating to the frequency increasing and reducing proton combinations. It is these chemical shifts and peak widths that were indicative of a molecules structure.

All NMR experiments were performed using a Varian 600 MHz NMR Spectrometer. 1D NMR was carried out using protein from a standard growth and cell harvest as detailed previously. The sample was concentrated to the desired concentration and contained 10 % $^{2}H_2O$ in a final volume of 350 µL into a glass NMR sample tube.

2D NMR was performed using protein obtained from a labelled growth and cell harvest as previously described. Similar to the 1D NMR, the sample was concentrated to the desired concentration and contained 10 % $^{2}H_2O$ in a final volume of 350 µL, but differing to the 1D sample preparation in that the sample was placed into a Shigemi tube.

2.4.3 Small Angle X-ray Scattering (SAXS)

SAXS is a technique used to obtain low-resolution structural data of macromolecules in solution. Resolution is approximately 13 Å due to the molecules within the solution constantly tumbling which results in an average scattering pattern. Theoretically SAXS could be used to produce higher angle diffraction generating higher-resolution data, however interpretation of this data would be difficult due to the fast decay rate of the X-rays combined with high levels of experimental noise. Despite its low resolution, SAXS experiments can be performed on macromolecules of larger size and increased levels of disorder or flexibility, which were often limiting factors in more high resolution techniques such as NMR and X-ray crystallography.

During SAXS analysis a collimated X-ray beam is fired through a mono-dispersed sample. Collisions with the sample cause the X-rays to scattered onto 2-dimentional (2D) detector (Figure 2.5). The intensities of each scattered beam, I(s), were radially averaged and its log is plotted against the scattering vector (s), which is the distance between the detection spot

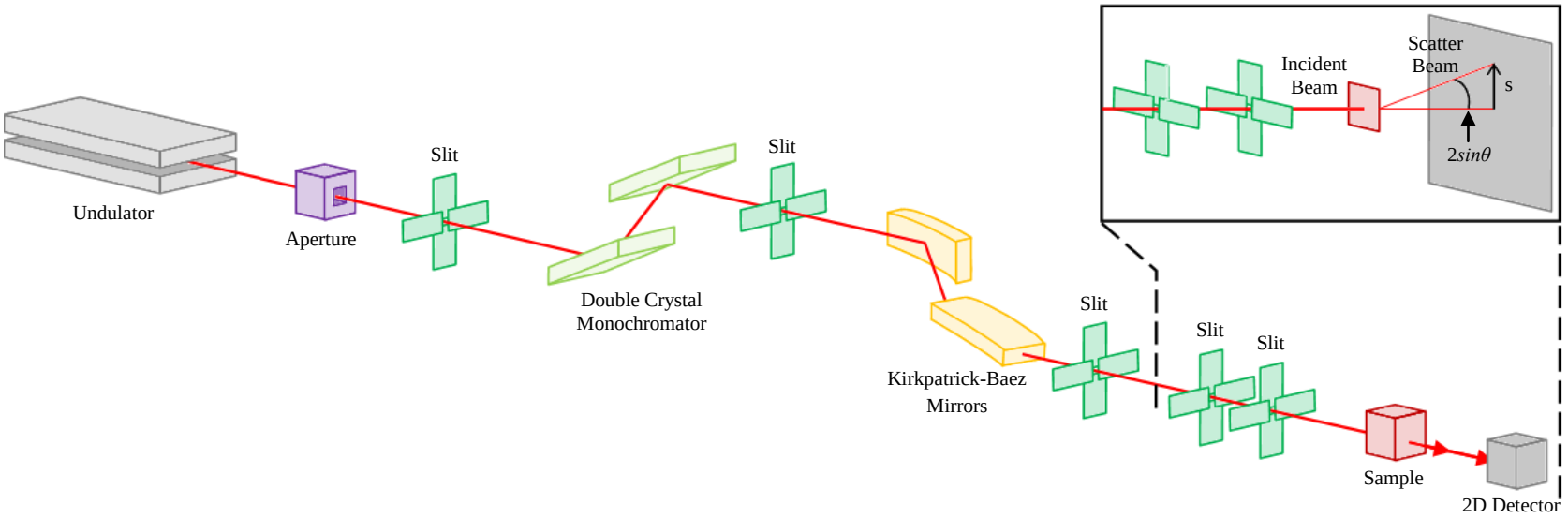


Figure 2.5 **Beamline I22 at the Diamond Light Source**. Adapted from the Diamond website, the schematic shows the undulator at the far left, where the X-rays are generated. The aperture and various slits serve as collimation pinholes, while the double crystal monochromator removes any extraneous wavelengths from the incident beam. The Kirkpatrick-Baez Mirrors focus the X-rays prior to the final collimation, before passing through the sample and onto the 2D detector.

of the unscattered X-ray (incident) beam and the detection spot of the scattered X-ray beam on the 2D detector, represented by the equation:

$$s = \frac{2sin\theta}{\lambda}$$

Where $2sin\theta$ is the angle of scattered X-ray and λ is the wavelength (0.01 to 10 nm). The low resolution of SAXS produces high intensities at small-angle and low intensities and large-angle diffraction. Isotropic scattering intensities at the small-angle region were given as I(Q) and defined by the following equation:

$$I(Q) = I(0)exp\left(\frac{-{R_g}^2Q^2}{3}\right)$$

Where $I(0)$ is the intensity at 0 scattering angle, R_g is the radius of hydration, and Q is the momentum transfer and is defined as $4\pi sin\theta/\lambda$ where 2θ and λ were the scattering angle and the X-ray wavelength respectively (Kamatari *et al.* 1998). This can then be used to elucidate information on the macromolecules structural-state (degree of globular shape) (Putnam *et al.* 2007), pair-distribution functions, P(r), maximum dimension (D_{max}), and real space Rg. Moreover, from this low resolution reconstructions can be made, aligned, averaged and subtracted for areas of low density to produce an average model.

SAXS experiments were carried out on the I22 beamline at the Diamond Light Source (Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire, England) with a wavelength (λ) of 1 angstrom (Å) at 12.4 kiloelectronvolts (keV), with a photon counting detector 3.25 m from the sample cell.

2.4.4 Dynamic Light Scattering (DLS)

DLS is commonly used to estimate a proteins' hydrodynamic radius (R_h) in solution, which is related to the diffusion constant (D) as described by the Stokes-Einstein equation:

$$R_h = \frac{kT}{6\pi\eta D}$$

Where *k is* the Boltzmann constant, *T* is temperature (K), and η is the solvent viscosity. R_h can also be used to estimate molecular weight (M_r), albeit roughly, as the shape of the molecule is not considered, using the following equation provided in the Dynamics version 5.26.60 software package (Protein Solutions):

$$M_r = (1.68 \times R_h)^{2.34}$$

DLS experiments were performed using a Protein Solutions DynaPro MSTC 800 (Wyatt) light scattering instrument. Protein solutions were allowed to equilibrate at the desired temperature prior to at least 20 measurements being taken to obtain the hydrodynamic radius (R_h).

2.5 Miscellaneous Protocols

2.5.1 Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Tuerk *et al.* 1990)

Materials

1. Ni-NTA resin (QIAGEN).
2. Centrifuge columns (Thermo Scientific).
3. 1 × PBS: 137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.4 mM K_2HPO_4, made with distilled water and autoclave sterilised[1] immediately.
4. 100 µM DNA library: A pool of oligonucleotides consisting of a random 30 nucleotide (nt) sequence flanked by know primer sequences:

   CCG CCG CAT AAT AT $(N)_{30}$ GCG CTA TAT AGG GCC

5. 1 M NaCl, made with distilled water and autoclave sterilised[1] immediately.
6. Microspin G-25 Desalting columns (GE Healthcare).

Method

All centrifugation was performed at 100 × g for 30 seconds unless otherwise stated. 200 µL Ni-NTA resin was applied to a centrifuge column, placed into a bullet tube and packed by centrifugation at 100 × g for 1 minute. The column was then washed with three consecutive centrifuge spins in 200 µL of 1 × PBS.

Then a pre-screened library was prepared by loading 100 µL of a 10 µM dilution of the random library in 1 × PBS onto the column and incubated at room temperature for 30 minutes. The column was then centrifuged in a fresh bullet tube and the flow-through (pre-screened library) was collected and put aside. Any bound DNA molecules were eluted after 5 minutes incubation with 200 µL of 1 M NaCl, centrifuged and discarded.

The column was then reconditioned back into 1 × PBS with three washes as described previously. His-tagged protein was immobilised by applying to the column, mixing by gentle agitation and incubating on ice for 20 minutes, any unbound protein was eluted by centrifugation and discarded. The immobilised protein was then screened against by adding 50 µL of the pre-screened library to the column and incubating on ice for 20 minutes before centrifuging and discarding the flow-though. Non-specifically bound DNA library molecules were removed with three washes of 1 × PBS performed as previously described. Any aptamers were eluted in by incubating the column in 100 µL 1 M NaCl for 5 minutes on ice, and then collected in a fresh bullet tube in the flow-though during centrifugation.

A Microspin G-25 Desalting column (GE Healthcare) was prepared by resuspending the resin by agitation and repacked by loosening the cap and removing the bottom closure prior to centrifugation for 1 minute at 735 × g. The flow-though was discarded and the column was placed into a new bullet tube. The aptamer sample was slowly applied to the desalting column then eluted by centrifugation a 735 × g for 2 minutes. The aptamer sample can be amplified using the polymerase chain reaction, or should be store at -20 °C until required.

2.5.2 Polymerase Chain Reaction (PCR)

Materials

1. 2 × GoTaq Master Mix (Promega).

2. Template DNA.
3. Forward and reverse primers:

Forward Primer: GGC GGC GTA TTA TA

Reverse Primer: GGC CCT ATA TAG CGC

100 µM made with distilled water.

Method

PCR reactions containing 1µM (0.5 µL of 100 µM) of each primer, the appropriate volume of template DNA, 25µl of 2 × GoTaq Master Mix, were made up to a final volume of 50 µL in distilled water and placed into a Robocycler Gradient 90 (Stratagene) PCR machine. PCR reactions were performed with an initial 95 °C denaturation for 1 minute followed by 35 cycles consisting of a 95 °C denaturing step for 30 seconds, a 60 °C annealing step for 30 seconds, and a 72 °C elongating step for 30 seconds.

2.5.3 pGEM-T Easy Vector (Promega) Cloning of PCR Products

Materials

1. 2 × Rapid Ligation Buffer (Promega).
2. Linear pGEM-T Easy Vector (50 ng μL^{-1}) (Figure 2.6).
3. DNA inserts.
4. T4 DNA Ligase (3 U μL^{-1})
5. LB agar (Lennox L Agar) plates mix (Fisher Scientific): 1 % w/v Tryptone, 0.5 % w/v yeast extract, 5 % w/v NaCl, 1 % w/v agar, made with distilled water and autoclave sterilised[1] immediately..
6. IPTG: 1M made with distilled water and filter sterilised[1] immediately.
7. X-gal: 20 mg mL^{-1} made with 100 % dimethylformaldehyde (DMF).

Method

Ligation reactions were prepared in bullet tubes containing 5 µL of 2 × Rapid Ligation Buffer, 1 µL of pGEM-T Easy Vector, 1 µL of T4 DNA Ligase, the appropriate volume of insert DNA, and made to a final volume of 10 µL with distilled water. Typically a 3:1 insert to vector molar ratio was used and can be calculated by:

$$\frac{ng\ of\ vector \times insert\ size\ (kbp)}{insert\ size\ (kbp)} \times insert{:}\ vector\ molar\ ratio = ng\ of\ insert$$

The ligation was performed at room temperature for 1 hour before being transformed (2.2.3) into *E. coli* Dh5α (Invitrogen) cells (Table 2.2) and grown at 37 °C overnight on LB agar plates supplemented with 100 µg mL^{-1} ampicillin, 1 mM IPTG, and 20 µg mL^{-1} X-gal. White colonies signify the successful incorporation of the insert, interrupting the *lacZ* gene and preventing the formation of functional β-galactosidase.

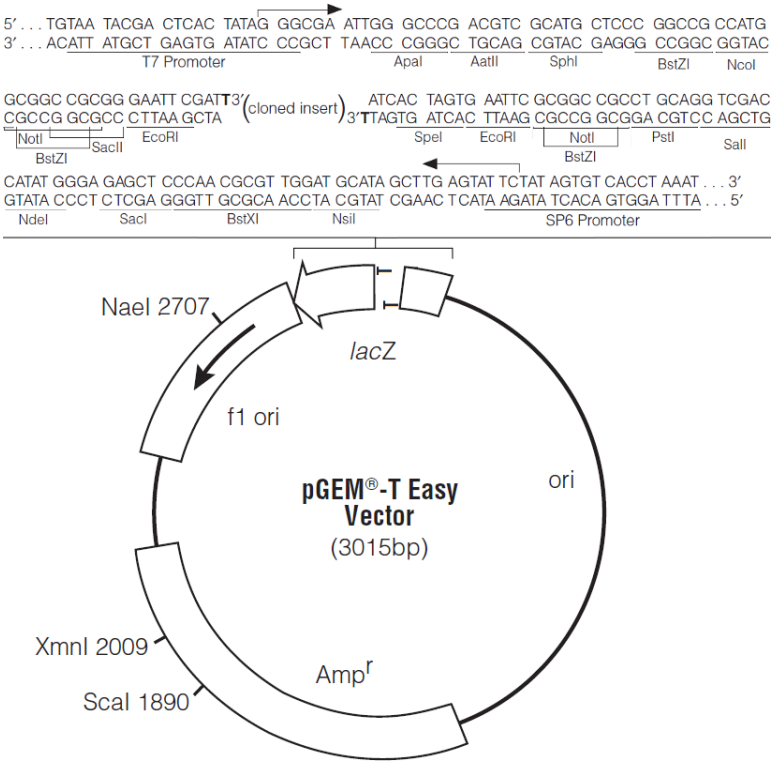


Figure 2.6 **pGEM-T Easy Vector Map.** Adapted from the Promega pGEM-T and pGEM-T Easy Vector Systems manual, the multiple cloning site (MCS) is highlighted at the top and details the associated restriction sites and the T-overhang cloning system. The MCS is under the control of a T7 *lac* promoter in the 5' direction and the SP6 promoter in the 3' direction.

2.5.4 Dot Blotting

Materials

1. TBST: 50 mM TRIS·HCl pH 8.0, 150 mM NaCl, 0.1 % Tween, made up to the desired volume with distilled water and filter sterilised[1] immediately.
2. Bovine serum albumin (BSA) (Fisher Scientific): 10 mg mL^{-1} made with distilled water and stored at -20 °C.
3. Two-part developing solution: Part 1; 2.5 mM luminol, 0.5 mM p-coumaric acid, 100 mM TRIS·HCl pH 8.0, made up to the desire volume in distilled water. Part 2; 0.05 % v/v hydrogen peroxide, 100 mM TRIS·HCl pH 8.0, made up to the desire volume in distilled water.

Method

Protein was immobilised to a nitrocellulose membrane by adsorption and drying for 15 minutes at room temperature. It was then incubated with a 1° probe in TBST with 5 % BSA at room temperature for 2 hours or 4 °C overnight. Next the membrane had three consecutive 5 minutes washes with TBST at room temperature before incubation with a 2° horseradish peroxidase (HRP) conjugated probe in TBST with 0.5 % BSA at room temperature for 1 hour. To finish, the membrane underwent a further three 5 minutes washes in TBST at room temperature. The blot was developed using a two-part developing solution and incubating for 1 minute before detecting with a Fujifilm LAS-3000 refrigerated camera at -20 °C under a 1 minute exposure time.

2.5.5 Drug Adduct Preparation

Materials

1. GWL-78 in 50% v/v methanol.
2. Duplex DNA Substrate.
3. Distilled water.

Method

The drug was incubated with the DNA substrate at the desired molar ratios at room temperature for 6 hours then transferred to -20 °C overnight.

2.6 Footnotes

2.6.1 [1]Sterilisation

Autoclave sterilisation was performed in a Midas 40 Priorclave autoclave at 121 °C and 100 kPa for 20 minutes. Plastic ware was then dried in a 65 °C oven for at least 24 hours. Filter sterilisation was performed by flowing the solution through a 0.2 µm filter membrane (Millipore) by means of either a vacuum pump (large volumes) or a syringe (small volumes). Any glassware intended for culture growths were autoclave sterilised, prior to making any media, containing only distilled water to remove residual detergents that can inhibit cell growth and transformation.

2.6.2 [2]Cell Strain Maintenance

All strains were stored at -70 °C. Every six months cells were maintained by growing 20 mL overnight cultures in Lysogeny Broth (LB) media. The following morning the cells were washed and stored in 1 freezing medium (36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 0.4 mM $MgSO_4 \cdot 7H_2O$, 1.7 mM $Na_3C_6H_5O_7$, 6.8 mM $(NH_4)_2SO_4$, 4.4% (v/v) glycerol).

CHAPTER THREE – PROTEIN PURIFICATION

3.1 Introduction

With approximately 60 % of the hypothetical proteins within the *P. falciparum* genome still to be annotated (Birkholtz *et al.* 2008) the lack of biological knowledge of the malaria parasite is abundantly apparent. Despite bioinformatic tools lending some idea of biochemical and structural properties to hypothetical proteins (Tuteja *et al.* 2006; Tuteja 2009; Tuteja 2010), this information can only be confidently revealed from biological experimentation which relies on the isolation of relatively large amounts of homogenous protein.

It is generally recognised that isolation of sufficient quantities of protein from native cell lines is unusual. Furthermore, the inherent risk associated with working, in this case, with *Plasmodium* species is relatively great. Therefore, the most common approach to purifying these proteins for studying is via heterologous expression in a host cell line such as *Escherichia coli*.

This chapter will describe the adaptations made to the original cloning and purification strategies for PFH45 (Pradhan *et al.* 2007) and PFDH60 (Pradhan *et al.* 2005) and the subsequent investigation into their storage conditions.

3.2 PFH45

PFH45 was one of the helicases proposed as potential drug-targets (Tuteja 2007) due to its origin from the most virulent malaria parasite, *Plasmodium falciparum*. PFH45 had been previously purified, and some characterisation had been undertaken (Pradhan *et al.* 2007; Pradhan *et al.* 2008). Due to the relative ease of the purification strategy to produce protein that displayed helicase activity, a clone for the His-tagged recombinant of PFH45 used in the aforementioned studies was obtained from Dr Renu Tuteja (Malaria Group,

International Centre for Genetic Engineering and Biotechnology, New Deli, India) with the intention to explore drug-target interactions using single-molecule studies.

3.2.1 Clone Confirmation by Restriction Digest

The pET-28a/*PFH45* clone, expressing a hexa-histidine (His_6) fusion of PFH45 with a thrombin (T) cleavage site (His_6-T-PFH45), obtained from Dr Renu Tuteja was digested with restriction enzymes BamHI and NotI to confirm the expected gene size of 1197 bp (Figure 3.1).

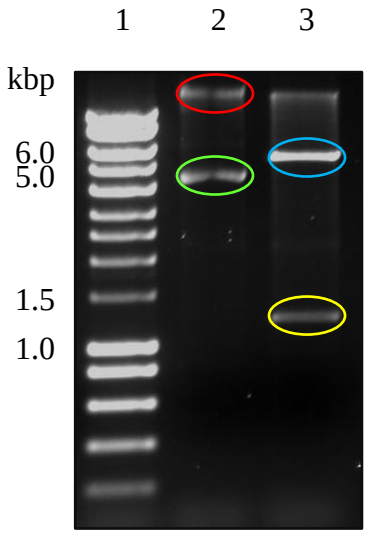


Figure 3.1 **Restriction Digest of the pET-28a(+)/*PfH45* clone from R. Tuteja.** 0.8 % agarose gel showing the double restriction digest of pET-28a(+)/*PfH45* by BamHI and NotI. Lane 1 shows the HyperLadder I marker (Bioline), lane 2 shows the undigested 6566 bp recombinant plasmid; in open circular (red) and supercoiled (green) conformations, and lane 3 shows double digest products; the 5369 bp linerised plasmid (blue) and the 1197 bp *PfH45* gene (yellow).

3.2.2 Pradhan and Tuteja (2007) Purification

An initial purification trial of pET-28a/*PFH45* was performed following the published purification strategy (Pradhan *et al.* 2007). In short, a 37 °C overnight seed culture of BL21 (DE3) [pLysS] transformants in LB media containing 40 µg mL^{-1} kanamycin was used to inoculate a 500 mL LB media growth, also containing 40 µg mL^{-1} kanamycin. The cells

were grown and induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside **(**IPTG) at 37 °C, harvested by centrifugation at 4000 × g for 12 minutes, lysed by sonication, and purified by binding to a 5 mL His Trap FF column (GE Healthcare) in 20 mM TRIS (pH 7.5), 50 mM NaCl, and 1 mM DTT, and eluted with the addition of 200 mM imidazole. The eluted fractions were visualised by SDS PAGE (Figure 3.2).

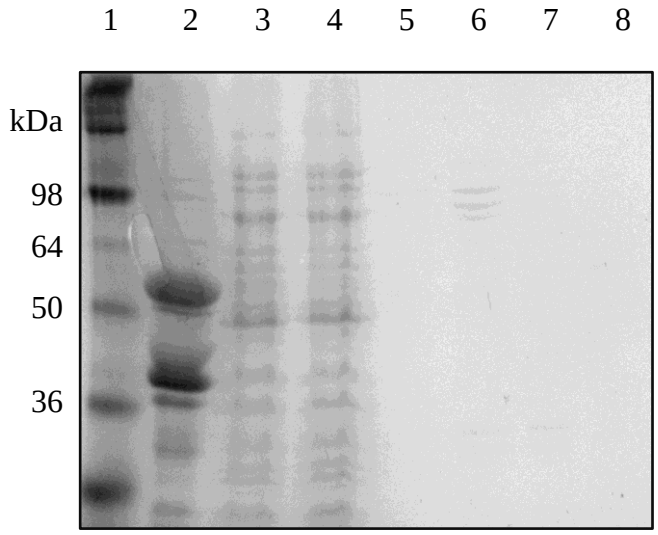


Figure 3.2 **Ni-column purification of Original His_6-T-PFH45.** 10 % SDS PAGE gel showing the flow-through, washed, and eluted fractions from a His Trap FF column (GE Healthcare) purification of His_6-Tagged PFH45. Lane 1 shows a SeeBlue Pre-stained Ladder (Invitrogen). Lane 2 shows the unbound flow-through, lanes 3 and 4 show the non-specifically bound fractions from the wash steps, and lanes 5 to 8 show the 200 mM Imidazole eluted fractions.

Due to the apparently low yield and multiple highly expressed bands in the flow-through, a more thorough protein expression analysis was performed. The uninduced and induced samples were analysed by Western blot with a monoclonal Anti-polyHistidine Peroxidise Conjugate (Sigma), which identified the presence of the His_6-T-PFH45 containing band (Figure 3.3).

3.2.3 Cloning of Codon-Optimised *PFH45*

As a result of the low yields and poor reproducibility of the purification strategy described by Pradhan *et al.* (2007), the *PFH45* gene was codon-optimised for *E. coli* codon bias in an

attempt to increase PFH45 expression (Fox *et al.* 2010). Codon-optimised *PFH45* was synthesised by GENEART (Regensburg, Bavaria, Germany), and cloned back into pET-28a(+) as well as the pGEX-6p-1 vector due to previous local successes.

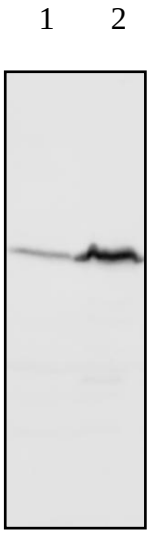


Figure 3.3 **Detection of His_6-T-PFH45 with an Anti-His Antibody.** Western blot analysis of the crude lysate from un-induced and induced pET-28a(+)/*PFH45* transformants in lanes 1 and 2 respectively.

3.2.4 Purification of Codon-Optimised GST-PP-PFH45

Codon-optimised pGEX-6p-1/*PFH45*, encoding a Glutathione S-tranferase (GST) fusion of PFH45 with a PreScission Protease (PP) cleavage site (GST-PP-PFH45), was transformed into BL21 (DE3) [pLysS] competent cells. A single colony was then used to spike a 37 °C overnight seed culture of LB media containing 100 µg mL^{-1} ampicillin which was in turn used to inoculate 500 mL LB media containing 100 µg mL^{-1} ampicillin. The cells were grown and induced with 1 mM IPTG at 37 °C, harvested by centrifugation at 4000 × g for 12 minutes, lysed by sonication and purified as described in the glutathione affinity protocol. The eluted fractions were visualised by SDS PAGE (Figure 3.4).

Like the original pET-28a(+) clone, there were multiple highly expressed bands that were difficult to determine the identity of. The first approach to overcome the low purity was to increase the overall expression of GST-PP-PFH45. Optimising the growth conditions by adjusting IPTG concentration, time and temperature proved ineffective. Therefore an alternative protocol exploiting a lactose-sucrose auto-induction was employed.

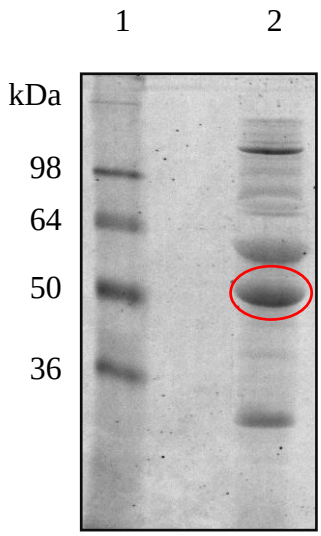


Figure 3.4 **GST-column purification of GST-PP-PFH45.** 10 % SDS PAGE gel showing the purified product obtained from an ÄKTA purification of GST-tagged PFH45. Lane 1 shows a SeeBlue Pre-stained Ladder (Invitrogen) while lane 2 shows the glutathione eluted protein. The eluted moiety shows the cleaved PFH45 circled in red.

Auto-induction was performed as a 6 hours 37 °C incubation of pGEX-6p-1/*PFH45* in ZYP-0.G media with 100 µg mL^{-1} ampicillin was used to inoculate 400 mL ZYP-5052 media containing 100 µg mL^{-1} ampicillin and grown overnight at 20 °C. The cells were harvested by centrifugation at 4000 × g for 12 minutes, lysed by sonication and purified as described above. Although greatly increasing the yield of GST-PP-PFH45, the majority of the fusion protein was insoluble and centrifuged into the pellet prior to the sample being loaded onto the column.

Further changes to lactose, sucrose, and glucose concentration, as well as time and induction temperature failed to improve solubility. Instead, the insoluble GST-PP-PFH45 was resolubilised in sarkosyl. This consisted of resuspending and sonicating the insoluble pellet obtained from centrifugation of the bacterial lysate in 5 mL 10 % sarkosyl. The sonicate was then centrifuged at 50000 × g for 30 minutes, the supernatant kept, diluted to 2 % sarkosyl, and made to 4 % triton X-100 and 30 mM CHAPS. This was used as the analyte for the purification protocol and the eluted fractions were visualised on a 10 % SDS PAGE gel (Figure 3.5).

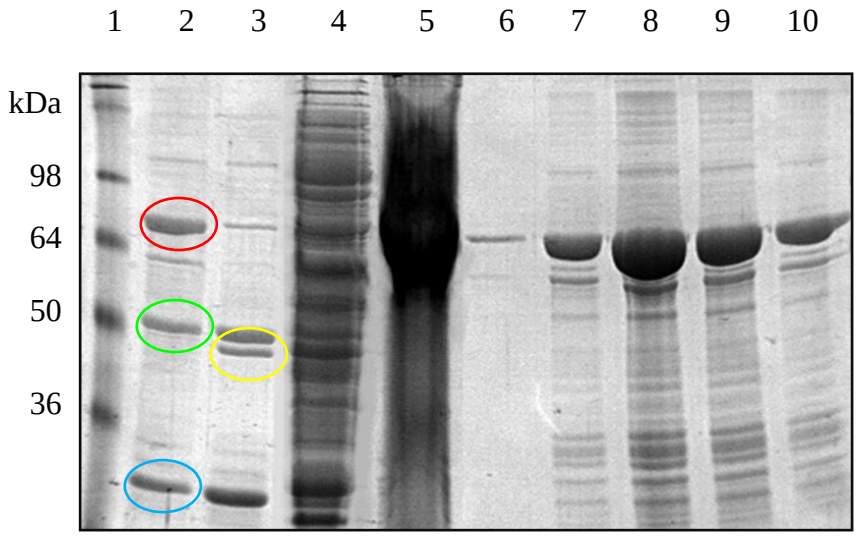


Figure 3.5 **Re-solubilised GST-PP-PFH45.** 10 % SDS PAGE gel showing SeeBlue Pre-Stained protein ladder (Invitrogen) in lane 1, and the purified GST-PP-PFH45 from Figure 3.3 in lane 2. Lane 2 has three highlighted bands; circled in red was the GST-PP-PFH45 fusion protein, while the further two bands have likely resulted from the degradation of the fusion protein into the PFH45 (green) and GST-PP moieties (blue). Lane 3 shows the sample after digestion with PreScission protease (yellow). Lanes 4 and 5 show the soluble and resolubilised fractions from the auto-induction respectively, while lanes 6 to 10 show the eluted fractions containing GST-PP-PFH45.

An attempt to further purify the protein product was carried out including DEAE cation-exchange chromatography, ammonium sulphate fractionation, and size exclusion. However there was no observable resolution of the contaminants (not shown).

The resolubilisation with sarkosyl greatly increased the yield; however the protein product had an optimum solubility of ~1 mg mL^{-1}. Furthermore, screening for helicase activity using a helicase substrate specificity screen, as detailed in chapter four, failed to produce any observable activity. Due to the limited solubility and no observable helicase activity the focus was again shifted, this time to the codon-optimised pET-28a(+) clone.

3.2.5 Purification of Codon-Optimised His_6-T-PFH45

Purification of the His_6-T-PFH45 fusion protein expressed from the codon-optimised pET-28a(+) clone was initially approached with both an auto induction protocol, as detailed above for the pGEX-6p-1 clone, and an IPTG induction as per the original purification

strategy described by Pradhan and Tuteja (2007). Varying temperatures and levels of induction, and the use of different growth media had little effect on the overall expression. It was also noted that His_6-T-PFH45 was expressed in small amounts prior to IPTG induction. The addition of 1 % glucose in the 37 °C overnight seed culture and 0.5 % in the growth was used to minimise any affects of leaky expression from the pET-28a vector (Loomis *et al.* 1967). The yield was largely unaffected by the presence of glucose, but a final protocol was established.

In short, a 37 °C growth in 400 mL 2 × YT and a 1 mM IPTG induction were used. However, unacceptable levels of contaminants were being co-purified. The 5 mL His Trap FF column (GE Healthcare) was replaced with a 10 mL Ni-NTA (Hochuli 1990) Superflow column (Qiagen) and additional steps to the purification protocol were explored. Varying imidazole concentrations revealed that lower concentrations had little observable affect on resolving His_6-T-PFH45 from the column contaminants, whereas 200 mM imidazole assured total elution of the His_6-T-PFH45 fusion protein. The introduction of a NaCl gradient appeared to have no observable affect, however a pH gradient from pH 7.5 down to 4.0 in Na_2HPO_4 buffer showed substantial reduction of the contaminants (Figure 3.6).

It was calculated from the pH gradient that some of His_6-T-PFH45 fusion began to dissociate from the column at a pH of less than 6.0. Therefore, subsequent purifications incorporated a pH 6.0 wash and the bound His_6-T-PFH45 was eluted in 200 mM imidazole. Protein product achieved from the Nickel affinity chromatography was further purified using size exclusion chromatography. This yielded a final His_6-T-PFH45 product (Figure 3.7) which showed good helicase activity at a 10:1 ratio of helicase to substrate.

Although a vast improvement on the overall purity of the PFH45 fusion protein, several contaminants still persisted. With resolving these contaminants in mind, samples of the protein product were applied to either DEAE or heparin columns. Although the DEAE made no noticeable difference, the heparin column did appear to resolve some of the contaminants. It was therefore decided to include heparin ion exchange chromatography after the nickel affinity chromatography but before the size exclusion chromatography. The heparin analyte carried forward from the nickel affinity chromatography step was dialysed against HEPES buffer A used in the heparin affinity protocol. The His_6-T-PFH45 moiety

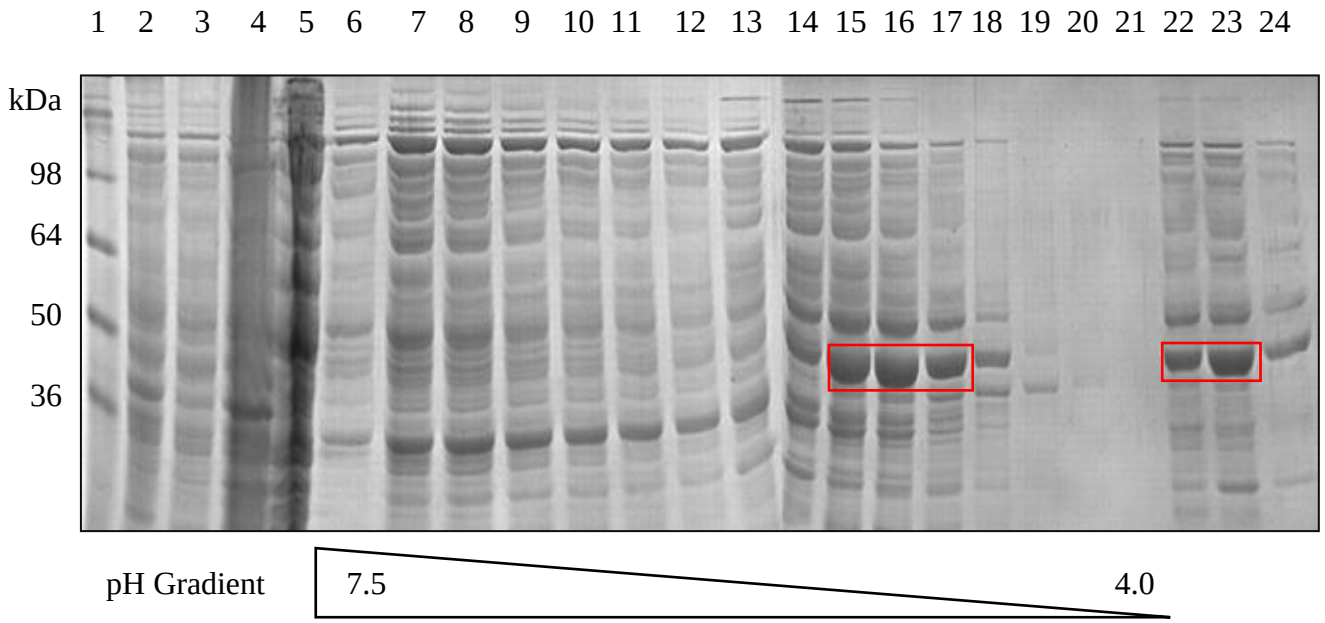


Figure 3.6 **pH Gradient across His_6-T-PFH45 Bound Ni-NTA Column.** 10 % SDS PAGE gel showing the SeeBlue Pre-Stained protein ladder (Invitrogen) in lane 1, the total cell lysate in lane 2, and the soluble and insoluble lysate fractions in lanes 3 and 4 respectively. Fractions collected during the pH gradient are shown in lanes 6 to 21, with lane 6 at pH 7.5 and lane 21 at pH 4.0. Lanes 22 to 24 show the fractions containing the residual bound protein eluted with 200 mM imidazole. The bands boxed in red are those that correspond to His_6-T-PFH45 based on size and expression.

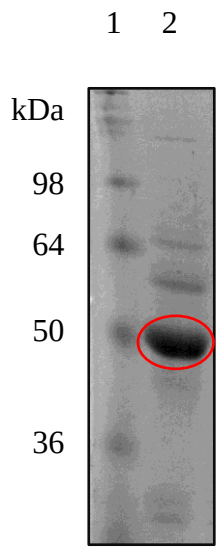


Figure 3.7 **Purification with a pH 6.5 wash and Size Exclusion Chromatography of His_6-T-PFH45.** 10 % SDS PAGE gel showing the SeeBlue Pre-Stained protein ladder (Invitrogen) in lane 1, while lane 2 shows the protein product obtained first from nickel affinity chromatography incorporating a pH 6.5 wash and eluted with 200 mM imidazole, then passed through a HiLoad 26/60 Superdex S200 Prep Grad column (GE healthcare). His_6-T-PFH45 was circled in red.

of the analyte remained in the mobile phase and was collected in the flow through, while the contaminants bound to the column (Figure 3.8).

Purified PFH45 samples from a heparin treated and a non-heparin treated preparation were screened for helicase activity. The heparin treated sample showed a marked decrease in helicase activity relative to that of the non-heparin treated sample. Succeeding purifications performed in the same way as described above for the heparin treated sample also showed this trend. Therefore it was decided to screen the contaminants from the heparin column for helicase activity. The predominant contaminants were separated using size exclusion chromatography (Figure 3.9). When each purified protein was screened for helicase activity, the higher molecular weight proteins (~90 kDa) conferred activity, whereas the lower molecular weight protein (~55 kDa) did not (not shown). Therefore it would seem that the removal of the contaminants during the heparin column treatment (the 90 kDa proteins) may be responsible for the decrease in helicase activity. It may also be the case that the two 90 kDa bands were the same protein, eluted from the size exclusion column in different oligomeric states.

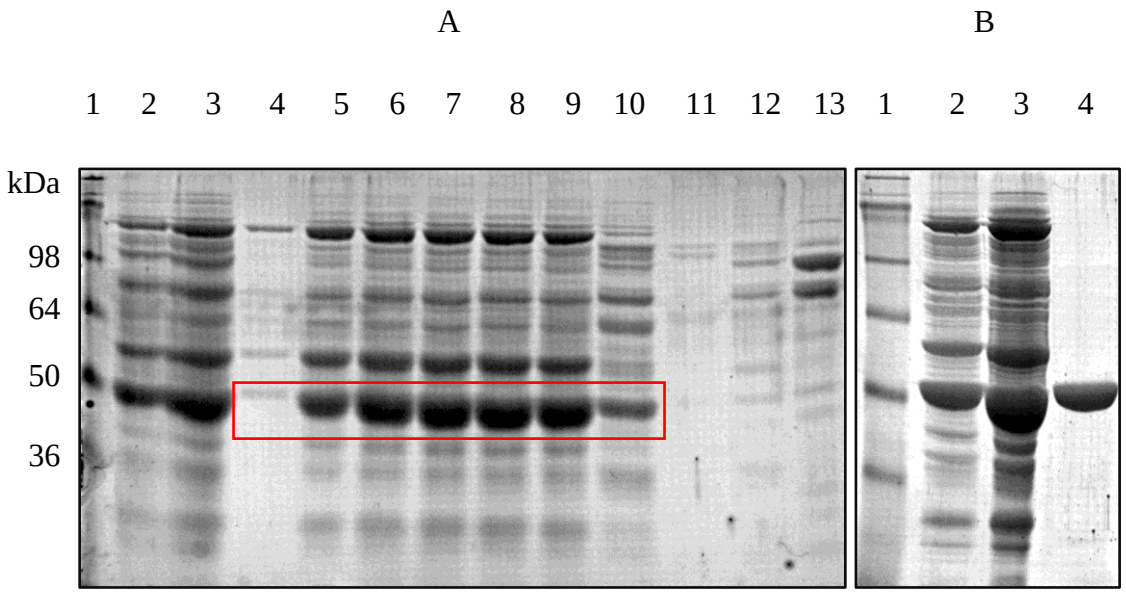


Figure 3.8 **Heparin Column Chromatography of Nickel Column Protein Product.** 10 % SDS PAGE gel A shows the SeeBlue Pre-Stained protein ladder (Invitrogen) in lane 1, while lane 2 shows the protein product obtained from nickel affinity chromatography and lane 3 shows it concentrated prior to loading onto the heparin column. Lanes 4 to 10 show the heparin flow through fractions containing the His_6-T-PFH45 fusion protein (red), and lanes 11 to 13 show the heparin eluted contaminants. 10 % SDS PAGE gel B shows the SeeBlue Pre-Stained protein ladder (Invitrogen) in lane 1, the protein product obtained from heparin ion exchange chromatography in lane 2 and again after being concentrated prior to loading onto the size exclusion column in lane 3. Lastly, the purified His_6-T-PFH45 obtained from the size exclusion chromatography was shown in lane 4.

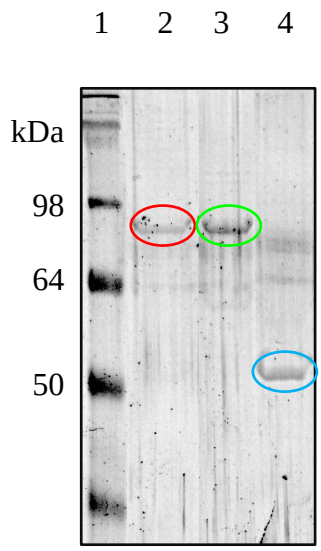


Figure 3.9 **Size Exclusion Chromatography of the Heparin Column Bound Phase during His_6-T-PFH45 Purification.** 10 % SDS PAGE gel showing the SeeBlue Pre-Stained protein ladder (Invitrogen) in lane 1, while lanes 2, 3 and 4 show the isolated dominant contaminants from the bound phase of the heparin chromatography step. Both the bands circled in red and green exhibit helicase activity; however the band circled in blue does not.

Western blot analysis was carried out using the same monoclonal Anti-polyHistidine Peroxidise Conjugate (Sigma) as described earlier (3.2.2.) to trace the presence of His-tagged PFH45 during each of the purification steps employed (Figure 3.10).

3.2.6 Final Purification Strategy of PFH45

The final growth and purification strategy (Figure 3.11) yielding ~12.5 mg of >95 % homogenous His_6-T-PFH45 per litre of growth media (~4.2 mg g^{-1} of wet weight of cells), based on a molar extinction coefficient of 29715 M^{-1} cm^{-1} and molecular weight of 48711.8 calculated using ProtParam (Wilkins *et al.* 1999), which exhibited helicase activity and had an A_{260}:A_{280} ratio of <0.6 was as follows.

The recombinant pET-27a(+) plasmid harbouring the codon-optimised *PFH45* gene for expression in *E. coli* was transformed into BL21 (DE3) [pLysS] *E. coli* cells (Stratagene). Bacterial cultures were grown in 400 mL of 2 × YT to an OD_{600} ~0.6 at 37 °C, at which

point the cells were induced with 0.5 mM IPTG and the growth continued for 3 hours at 37 °C. The cells were harvested by centrifugation at 4000 × g for 15 minutes at 4 °C, and stored at -20 °C.

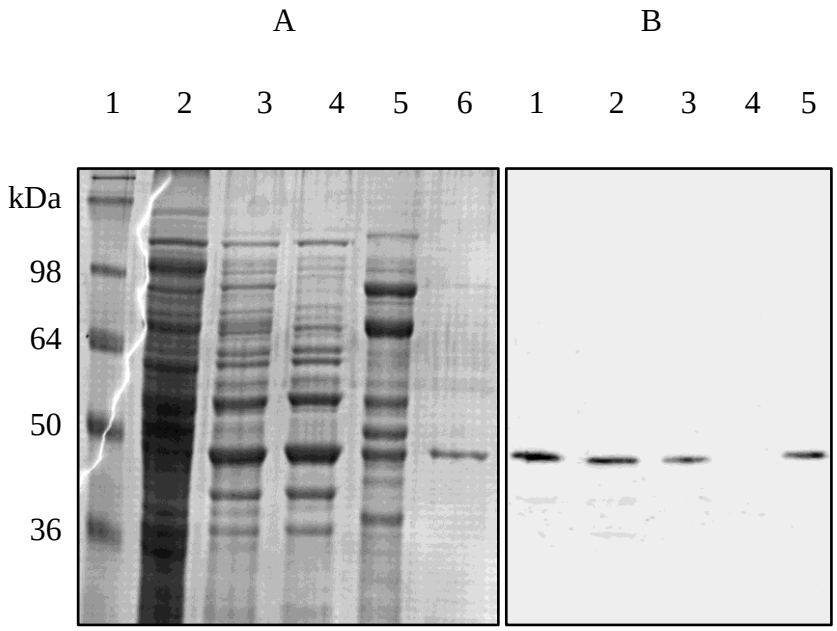


Figure 3.10 **Western Blot Tracing His_6-T-PFH45 throughout its Purification.** 10 % SDS PAGE gel A shows a SeeBlue Pre-Stained protein ladder (Invitrogen) in lane 1, and lane 2 shows the cell lysate. Lane 3 shows the eluant from the Ni-NTA column. Lane 4 shows the flow through of the heparin column while lane 5 shows the bound phase eluted with high salt. Lane 6 shows the pure His6-T-PFH45 obtained from size exclusion chromatography. Panel B shows the nitrocellulose membrane with the electro-transferred samples representative of lanes 2 to 6 shown on gel A, western blotted against with monoclonal Anti-polyHistidine Peroxidise Conjugate (Sigma).

The cell pellet from the 400 mL growth was thawed, resuspended and lysed in 1 × PBS, 1 mM DTT, 1 mM Benzamidine, and 0.1 mM PMSF by sonication (40 % amplitude at a max temperature of 10 °C for a duration of 3.3 seconds on and 9.9 seconds off for a total time of 20 minutes). The sonicate was clarified by centrifugation at 50000 × g for 1 hour at 4 °C.

The resulting supernatant was loaded onto a 10 mL Ni-NTA agarose resin (Qiagen) column pre-equilibrated in disodium phosphate buffer (50 mM Na_2HPO_4 pH 7.5, 500 mM NaCl, 1 mM DTT, and 10 % glycerol). The column was washed with three column volumes of disodium phosphate buffer, followed by three column volumes of modified pH

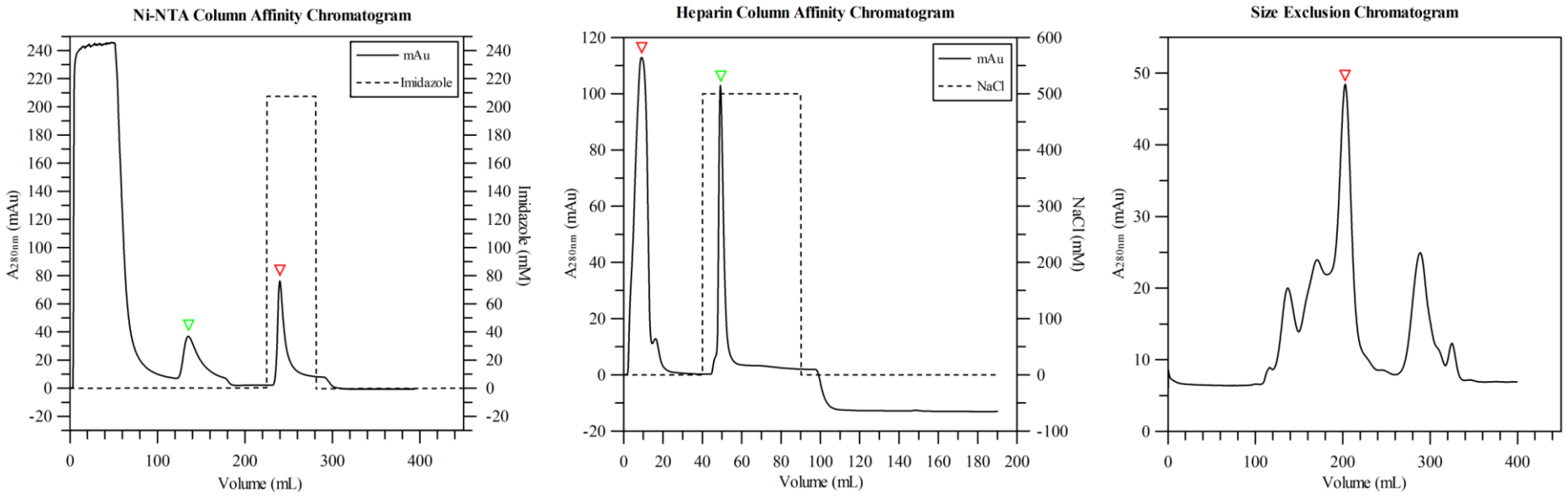


Figure 3.11 **Chromatagrams Obtained from the Three-Stage Purification Strategy of His_6-T-PFH45.** In the first of three purification protocols, the clarified supernatant obtained from the centrifugation of the bacterial lysate was injected onto a Ni-NTA agarose resin (Qiagen) column. Non-specifically bound contaminants were removed with a pH 6.0 wash (green arrow). The His_6-T-PFH45 was eluted from the bound phase with 200 mM imidazole (red arrow). The eluate from the nickel column was buffer exchanged for a NaCl concentration of <100 mM and then injected onto a HiTrap Heparin HP column (GE Healthcare) where the His_6-T-PFH45 remained into the mobile phase (red arrow). The heparin bound contaminants were eluted in 500 mM NaCl (green arrow). The His_6-T-PFH45 containing flow through fractions were concentrated down to a suitable volume (2 mL) and injected onto a HiLoad 26/60 Superdex S200 Prep Grad column (GE Healthcare) where the pure His_6-T-PFH45 eluted at approximately 200 mL (red arrow).

disodium phosphate buffer (50 mM Na_2HPO_4 pH 6.0, 500 mM NaCl, 1 mM DTT, and 10 % glycerol). The bound phase was conditioned back into pH 7.5 with three column volumes of disodium phosphate buffer, and then eluted with 200 mM imidazole in HEPES buffer as described in the nickel affinity protocol.

The eluted sample was buffer exchanged into HEPES buffer A used in the heparin affinity protocol to achieve an overall salt concentration of 20 mM or more but less than 100 mM ($20 \leq NaCl < 100$ mM). The sample was further purified by applying the sample to a 5 mL HiTrap Heparin HP column (GE Healthcare) pre-equilibrated in HEPES buffer A.

Flow through from the column was then concentrated using a 30 kDa molecular weight cut-off (MWCO) centrifugal concentrator (GE Healthcare) and applied to a HiLoad 26/60 Superdex S200 Prep Grad column (GE Healthcare) in HEPES buffer used in the size exclusion protocol. Fractions containing pure His_6-T-PFH45 were concentrated to 50 µM and stored as 50 µL aliquots at -20 °C.

3.2.7 Thrombin Cleavage

Thrombin was used for an on-column (Ni-NTA) cleavage of His_6-T-PFH45 using a 1:2000, mass to mass ratio of thrombin to His_6-T-PFH45 at 20 °C for 16 hours. Thrombin was removed from the eluted PFH45-thrombin mixture using a benzamidine affinity column, yielding pure untagged PFH45 in the flow-through. There was no observable change in PFH45 activity as a result of the cleavage and due to the additional steps to the already convoluted purification strategy, His-tag cleavage was not performed prior to the biochemical analysis of PFH45.

3.2.8 Storage Conditions

Repeated helicase activity assays were performed over consecutive days using purified His_6-T-PFH45 stored at various temperatures and glycerol concentrations. Loss of activity was observed in protein stored at 4 °C regardless of the glycerol concentration, while proteins stored at -20 °C and subjected to multiple freeze-thaw cycles did not show a

significant loss of activity providing the storage buffer contained at least 10 % glycerol to serve as a cryoprotectant. Successive freeze-thaws of -20 °C stocks in 10 % glycerol did result in decrease in SDS PAGE running size of <5 kDa. This was determined to be the loss of the His-tag, illustrated by a lack of anti-His detection using Western blotting. Long term storage (greater than one week) at 4 °C of His_6-T-PFH45 resulted in the consistent degradation into two polypeptides of approximately 30 kDa and 25 kDa which could be separated by size exclusion (Figure 3.12).

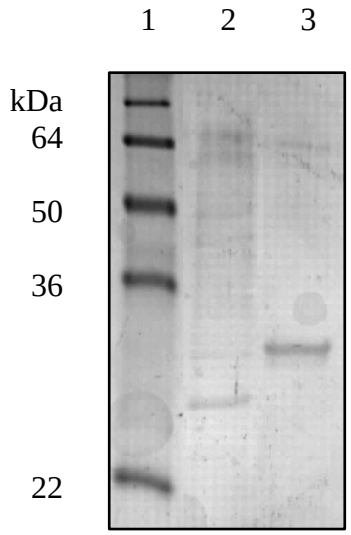


Figure 3.12 **Purified Products from the Consistent Degradation of PFH45.** 10 % SDS PAGE showing a SeeBlue Pre-Stained protein ladder (Invitrogen) in lane 1, the smaller (~25 kDa) of the two products in lane 2, and the larger (~30 kDa) of the two products in lane 3.

3.3 PFDH60

In addition to PFH45, Tuteja *et al.* (2007) also proposed PFDH60 as a novel drug-target. The 60 kDa truncated namesake of the protein has been previously purified (Pradhan *et al.* 2005; Pradhan *et al.* 2005) and a clone was obtained from Dr Renu Tuteja (Malaria Group, International Centre for Genetic Engineering and Biotechnology, New Deli, India). Again yields suitable for the proposed biochemical and structural experimentation were unattainable. As for *PFH45*, a codon-optimised version of *PFDH60* (GENEART, Regensburg, Bavaria, Germany) was cloned into pET-28a(+) and pGEX-6P-1. A single-step GST-fusion capture purification protocol failed to yield any discernible GST-PP-

PFDH60. Therefore due to the success of the codon-optimised clone of His_6-T-PFH45, the pET-28a(+) clone was chosen to pursue viable PFDH60.

3.3.1 Clone Confirmation by Restriction Digest

The codon-optimised pET-28a(+)/*PfDH60* clone, expressing a hexa-histidine (His_6) fusion of PFDH60 with a thrombin (T) cleavage site (His_6-T-PFDH60), was digested with restriction enzymes BamHI and NotI confirming the expected gene size of 2229 bp (Figure 3.13).

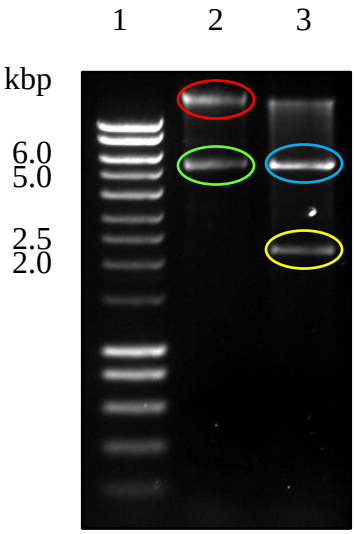


Figure 3.13 **Restriction Digest of the pET-28a(+)/*PfDH60* codon-optimised clone.** 0.8 % agarose gel showing the double restriction digest of pET-28a(+)/*PfDH60* by BamHI and NotI. Lane 1 shows the HyperLadder I marker (Bioline), lane 2 shows the undigested 7598 bp recombinant plasmid; in open circular (red) and linear (green) confirmations, and lane 3 shows double digest products; the 5369 bp linerised plasmid (blue) and the 2229 bp *PfDH60* gene (yellow).

3.3.2 Purification of Codon-Optimised His_6-T-PFDH60

Due to the difficulty originally experienced with PFH45, expression analysis was carried out following the induction of pET-28a(+)/*PfDH60* with various IPTG concentrations and at different temperatures. Expression was not significantly affected by IPTG concentration

or incubation temperature, and an auto-induction protocol had little effect; therefore further growths were induced with 1 mM IPTG at 37 °C for 3 hours.

Purification of the His_6-T-PFDH60 fusion protein expressed from the codon-optimised pET-28a(+) clone was performed as per the final purification strategy for His_6-T-PFH45, but with minor modifications. Like the His_6-T-PFH45 nickel affinity step of purification, several histidine-containing proteins were captured from the PFDH60 bacterial sonicate. Two approaches to eliminate this were explored, including the incorporation of either a pH wash or a routine imidazole wash. A pH gradient did not provide any resolution of the His_6-T-PFDH60 from the contaminants, but an isocratic gradient of imidazole showed removal of the vast majority of contaminants at 30 mM imidazole (Figure 3.14). Therefore, the nickel affinity step in future preparations included a 30 mM imidazole wash prior to eluting the His_6-T-PFDH60 fusion protein with 100 mM imidazole.

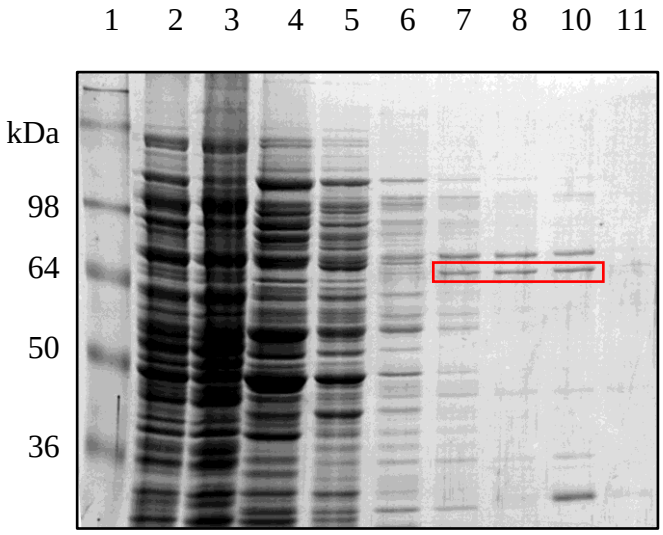


Figure 3.14 **Nickel Affinity Chromatography with an Isocratic Imidazole Gradient.** The 10 % SDS PAGE gel shows the soluble fraction of the lysate, flow-through, and various imidazole concentration eluted fractions from a Ni-NTA affinity column (GE Healthcare) purification of His_6-Tagged PFDH60. Lane 1 shows a SeeBlue Pre-stained Ladder (Invitrogen). Lane 2 shows the total soluble fraction from the cell pellet lysate, lane 3 shows the flow-through, and lanes 4 to 11 show the eluted bound phase fractions by 10, 20, 30, 40, 50, 100, and 200 mM imidazole respectively. The band corresponding in size (~ 90.4 kDa) to His_6-T-PFDH60 is present in the fractions eluted in 40 to 100 mM imidazole (red).

Based on the observation that the heparin affinity removed native *E. coli* helicases during the His_6-T-PFH45 purification, it was considered prudent to include it in the PFDH60 purification strategy. Similar to the PFH45 heparin affinity step, His_6-T-PFDH60 remained in the mobile phase and was collected in the flow-through while a ~90 kDa protein was eluted from the bound phase resembling that described in the PFH45 heparin affinity step.

Size exclusion chromatography was included as the final purification step. The corresponding His_6-T-PFDH60 band was traced during the described purification strategy by western blot analysis with the monoclonal Anti-polyHistidine Peroxidise Conjugate (Sigma) as described in 3.2.2 (Figure 3.15).

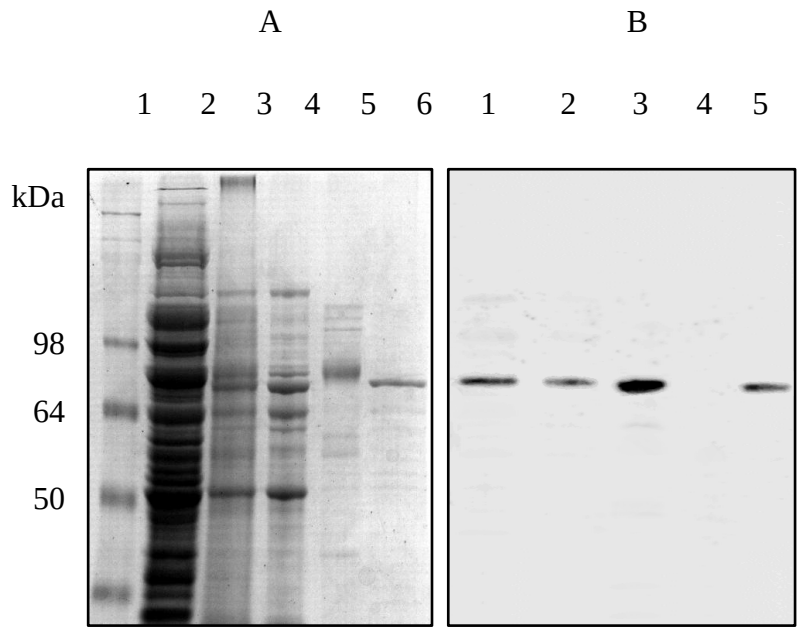


Figure 3.15 **Western Blot Tracing His_6-T-PFDH60 throughout its Purification.** 10 % SDS PAGE gel A shows a SeeBlue Pre-Stained protein ladder (Invitrogen) in lane 1, and lane 2 shows the cell lysate. Lane 3 shows the eluant from the Ni-NTA column. Lane 4 shows the flow through of the heparin column while lane 5 shows the bound phase eluted with high salt. Lane 6 shows the pure His6-T-PFH45 obtained from size exclusion chromatography. Panel B shows the nitrocellulose membrane with the electro-transferred samples representative of lanes 2 to 6 shown on gel A, western blotted against with monoclonal Anti-polyHistidine Peroxidise Conjugate (Sigma).

3.3.3 Final Purification Strategy of PFDH60

The final growth and purification strategy (Figure 3.16) yielding ~0.7 mg of >95 % homogenous His_6-T-PFDH60 per litre of growth media (~0.2 mg g^{-1} of wet weight of cells), based on a molar extinction coefficient of 80985 M^{-1} cm^{-1} and molecular weight of 90421.9 calculated using ProtParam (Wilkins *et al.* 1999), which exhibited helicase activity and had an A_{260}:A_{280} ratio of <0.6 was as follows.

The recombinant pET-27a(+) plasmid with the *PFDH60* gene was transformed into BL21 (DE3) [pLysS] *E. coli* (Stratagene). Bacterial cultures were grown in 400 mL of 2 × YT to an OD_{600} ~0.6 at 37 °C, at which point the cells were induced with 0.5 mM IPTG and the growth continued for 3 hours at 37 °C. The cells were harvested by centrifugation at 4000 × g for 15 minutes at 4 °C, and stored at -20 °C.

The cell pellet from the 400 mL growth was thawed, resuspended and lysed in 1 × PBS, 1 mM DTT, 1 mM Benzamidine, and 0.1 mM PMSF by sonication (40 % amplitude at a max temperature of 10 °C for a duration of 3.3 seconds on and 9.9 seconds off for a total time of 20 minutes). The sonicate was clarified by centrifugation at 50000 × g for 1 hour at 4 °C.

The supernatant was treated on a 10 mL Ni-NTA agarose resin (Qiagen) column as described in the nickel affinity protocol, with an additional three column volumes of 15 % HEPES buffer against disodium phosphate buffer to provide the 30 mM imidazole wash prior to eluting with 100 mM imidazole (50 % HEPES buffer against disodium phosphate buffer).

The eluted sample was buffer exchanged into HEPES buffer A used in the heparin affinity protocol to achieve an overall salt concentration 20 mM or more but less than 100 mM ($20 \leq NaCl < 100$ mM). The sample was further purified using a 5 mL HiTrap Heparin HP column (GE Healthcare) as described in the heparin affinity protocol.

Flow through from the column was then concentrated using a 50 kDa molecular weight cut-off (MWCO) centrifugal concentrator (GE Healthcare) and applied to a HiLoad 26/60 Superdex S200 Prep Grad column (GE Healthcare) as detailed in the size exclusion

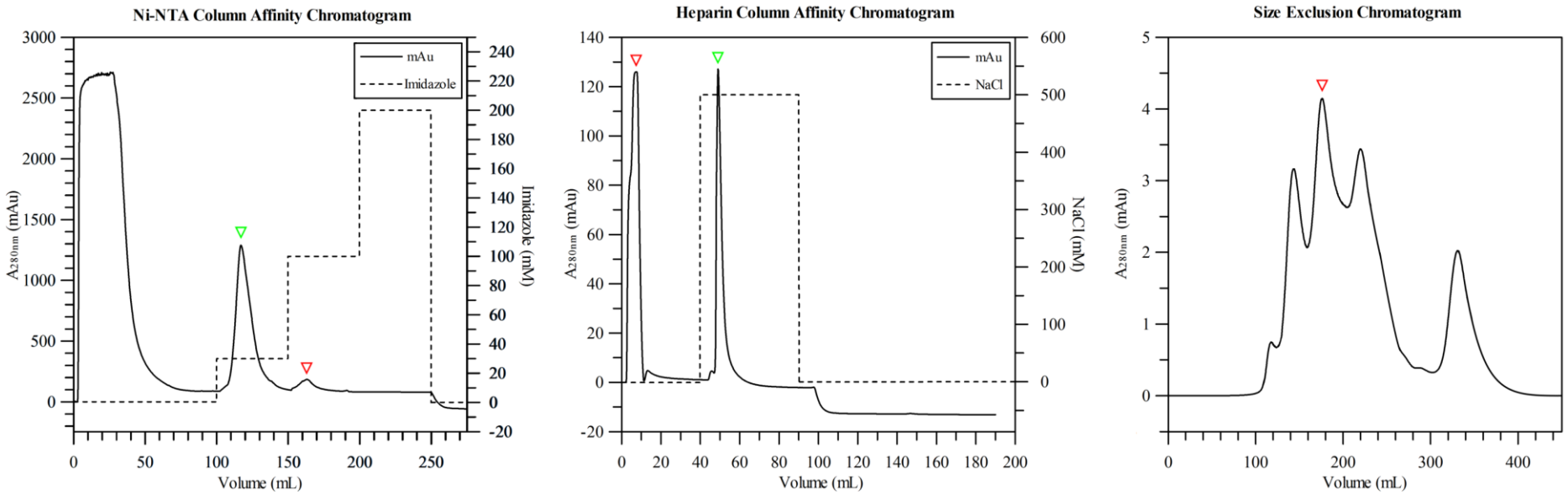


Figure 3.16 **Chromatagrams Obtained from the Three-Stage Purification Strategy of His_6-T-PFDH60.** In the first of three purification protocols, the clarified supernatant obtained from the centrifugation of the bacterial lysate was injected onto a Ni-NTA agarose resin (Qiagen) column. Non-specifically bound contaminants were removed with a 30 mM imidazole wash (green arrow). The His_6-T-PFDH60 was eluted from the bound phase with 100 mM imidazole (red arrow). The eluate from the nickel column was buffer exchanged for a NaCl concentration of <100 mM and then injected onto a HiTrap Heparin HP column (GE Healthcare) where the His_6-T-PFDH60 remained into the mobile phase (red arrow). The heparin bound contaminants were eluted in 500 mM NaCl (green arrow). The His_6-T-PFDH60 containing flow through fractions were concentrated down to a suitable volume (2 mL) and injected onto a HiLoad 26/60 Superdex S200 Prep Grad column (GE Healthcare) where the pure His_6-T-PFDH60 elutes at approximately 180 mL (red arrow).

protocol. Fractions containing pure His_6-T-PFDH60 were concentrated to 10 µM and stored at 4 °C.

3.3.4 Thrombin Cleavage

As for PFH45, an on-column (Ni-NTA) cleavage of His_6-T-PFDH60 using a 1:2000, mass to mass ratio of thrombin to His_6-T-PFDH60 was performed at 20 °C for 16 hours. Thrombin was removed from the eluted PFDH60-thrombin mixture using a benzamidine affinity column, yielding pure untagged PFDH60 in the flow-through. There was no observable change in PFDH60 activity resulting from cleavage. Therefore, as for PFH45 His-tag cleavage was not performed prior to the biochemical analysis of PFDH60.

3.3.5 Storage Conditions

Repeated helicase activity assays were performed over consecutive days using purified His_6-T-PFDH60 stored at various temperatures and glycerol concentrations. Loss of activity was observed after 1 week in protein stored at any temperature (4 to -20 °C) regardless of the glycerol concentration (0-50 % v/v).

3.4 Discussion

Due to the cooperative nature of helicase activity and their capability to bind nucleic acids, helicases can prove difficult to purify to a high degree of homogeneity, and *Plasmodium* proteins have been identified as being particularly difficult to purify from heterologous expression systems (Birkholtz *et al.* 2008). Additionally, most proteins benefit from quick and uncomplicated purification strategies to reduce susceptibility to proteolysis and denaturation. However, single column chromatography rarely yields proteins with greater than 90 % purity (Tahmaseb *et al.* 2010). There are at least 11 different DNA helicases in *Escherichia coli* (Matson *et al.* 1990; Matson 1991; Matson *et al.* 1994), and therefore the

production and purification of proteins using expression strains of *E. coli* means that there was always the possibility of contaminating native helicase being carried forward into the biochemical analysis, resulting in inaccurate results. This justified the more rigorous strategy of purification, not only to yield a more homogenous protein sample but also to minimise the potential of helicase contamination.

3.3.1 PFH45

Prior to the codon-optimisation, PFH45 expression was too poor for it to be detected in the eluent from nickel affinity chromatography. Use of the Sequence Manipulation Suite: Codon Usage tool (Stothard 2000) and ALIGN (Myers *et al.* 1988) on the native and codon optimisation sequences (Appendix 2), indicated that codon optimisation for expression in *E. coli* resulted in a change of approximately 26 % of the 1197 nucleotide native *P. falciparum* sequence, corresponding to 104 codon triplets (Appendix 3). This resulted in an apparent increase of expression which gave rise to usable yields for biochemical and structural analysis.

The lack of observable activity from the sarkosyl resolubilised GST-PP-PFH45 obtained from the codon-optimised pGEX-6p-1 clone may have been due to the protein being rendered inactive due to misfolding (Goldberg 2003), coinciding with its initial insolubility. It may also have been that the helicase was being inhibited by any sarkosyl carried through from the purification protocol (Kovelman *et al.* 1990). These observations justified moving onto the codon-optimised pET-28a(+) clone.

The pH wash incorporated in the His_6-T-PFH45 purification strategy did not appear to affect the helicase activity, which is detailed in chapter four. Even protein eluted after the pH 4.0 gradient showed activity, although it should be noted that exposure to the relatively low pH was not prolonged. Although not common, there have been reported cases of helicases remaining active across a broad pH range (Lam *et al.* 2004; Pradhan *et al.* 2005), lending credibility to pH adjustment as a purification tool.

3.3.2 PFDH60

Similar to observations on PFH45 expression prior to codon optimisation, PFDH60 expression in *E. coli* was equally poor and therefore difficult to detect in the eluent from nickel affinity chromatography. Again, analysis of PFDH60 native and codon optimised sequences (Appendix 4) using the Sequence Manipulation Suite: Codon Usage tool (Stothard 2000) and ALIGN (Myers *et al.* 1988) indicated that codon optimisation for expression in *E. coli* resulted in a change of approximately 23 % of the 2229 nucleotide native *P. falciparum* sequence, corresponding to 168 codon triplets (Appendix 5). Although codon optimisation did increase the yield, its effect was not as noticeable as it was for PFH45. It should be noted that expression of the native PFDH60 was poorer than that of the native PFH45, and so any increase in yield may be relative. However, the yield was still sufficient for further biochemical and structural analysis.

The difficulty in maintaining PFDH60 activity under different storage conditions is similar to that observed by Pradhan *et al*. (2007). As a result, subsequent biochemical and structural analysis of PFDH60 would have to be performed within three days of its purification.

CHAPTER FOUR: BIOCHEMICAL CHARACTERISATION

4.1 Introduction

Before studies into the exploitation of a protein for biotechnological or pharmaceutical applications can be undertaken, the specific activities of that protein need to be identified. The better understood a protein is, the easier it is to exploit it. A recent study on Human Immunodeficiency Virus (HIV) 1 reverse transcriptase (RT) detailing its inhibition mechanism (Das *et al.* 2012) would not have been possible had its activities not been investigated over the past 25 years (Tanese *et al.* 1986; Farmerie *et al.* 1987; Larder *et al.* 1987; Hizi *et al.* 1988; Mous *et al.* 1988; Müller *et al.* 1989). Although the availability and advancement of technologies have made the investigation of protein activities faster, this example demonstrates the level of biochemical characterisation that needs to be performed if a protein is to be exploited as, in the case of this study, a novel drug-target.

Singleton *et al.* (2007) described the biochemical properties of helicases in terms of rate, processivity, directionality, step size, and active or passive mechanisms, and that the classification nomenclature should be representative of a particular helicase activity. For example, a helicase superfamily one member which translocates in the 3'-5' direction should be classified as a SF1A helicase. Furthermore, if that same helicase were only capable of acting on a substrate containing a single-stranded nucleic acid region it would be referred to as a SF1Aα helicase. Due to the clarity and comprehensiveness of the Singleton *et al.* (2007) paper, it was decided that the biochemical characterisation in this study would incorporate experiments which could attempt to address each of these biochemical properties.

This chapter will describe these experiments, how they compare with the previous experimentation carried out by the Tuteja group, and how precisely PFH45 and PFDH60 can been described based on the Singleton *et al.* (2007) mode of classification.

4.1 PFH45

Although the intention was to explore drug-target interactions using single-molecule studies of PFH45, it was felt that due to the alternative purification strategy outlined in chapter three, it would be prudent to repeat and expand on the preliminary biochemical characterisation by the Tuteja group. The biochemical analysis carried out by the Tuteja group had concluded that PFH45 depends on ATP and Mg^{2+} cofactors for optimum helicase activity (Pradhan *et al.* 2007), but can utilise the hydrolysis of dATP, dCTP, and to a lesser extent UTP (Pradhan *et al.* 2008). Additionally, they had shown PFH45 to bind radiolabelled ATP and that it was capable of unwinding fork-like dsDNA and dsRNA in both the 5'-3' and the 3'-5' directions (Pradhan *et al.* 2007). They had also demonstrated that PFH45s ability to bind ATP resides in the N-terminal residues of the protein, while the C-terminal residues retain 75 and 78 % of the full-length proteins' propensity for RNA and DNA binding respectively, and that phosphorylation of its serine and threonine residues results in a two to three-fold up-regulation in helicase activity (Pradhan *et al.* 2008).

4.1.1 *In silico* Analyses

Prior to laboratory experimentation, PFH45 was scrutinised using a variety of *in silico* techniques. An initial search using the unspliced, spliced, and codon-optimised DNA sequences of PFH45 (Appendix 2) with the Basic Local Alignment Search Tool (BLAST) server (Altschul *et al.* 1990; Johnson *et al.* 2008) revealed sequence identity to numerous DEAD box helicases. Most notably similar was eukaryotic initiation factor (EIF) 4A from the closely related cattle parasitic protozoan *Theileria annulata* (taxon Class *Aconoidasida* (Mehlhorn *et al.* 1980)), as well as other EIF4A orthologues in different species. An alignment of the PFH45 protein sequence with that of human EIF4A1 (NCBI accession number: NP_001407), one biologically active form of EIF4As' three isoforms (Conroy *et al.* 1990; Yoder-Hill *et al.* 1993), using ALIGN (Myers *et al.* 1988) (Appendix 6) revealed a sequence identity of 66.7 %, with significant conservation of the core SF2 family motifs between the two sequences (Figure 4.1). These observations support those of Pradhan *et al.* (2007), while more recently Tuteja *et al.* (2010) have renamed PFH45 as PFEIF4A. For these reasons, this body of work will continue to refer to PFH45 as PFEIF4A.

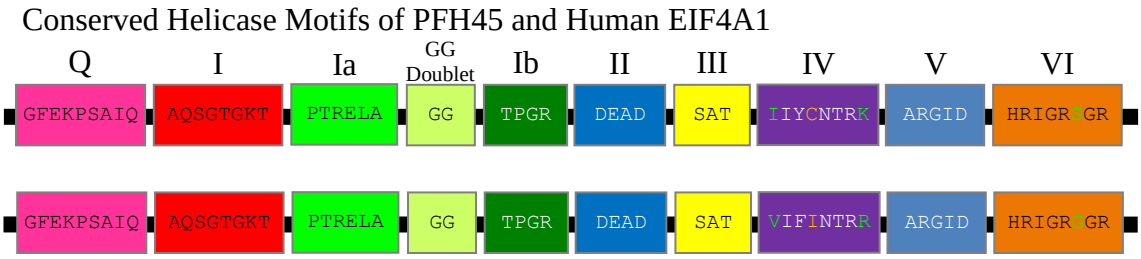


Figure 4.1 **Schematic Based on ALIGN results of PFH45 and EIF4A1.** Each diagram represents the SF2 motifs present in PFH45 (top) and EIF4A1 (bottom), showing the N-termini on the left and C-termini on the right. Open boxes represent the conserved helicase motifs, and letters inside the boxes are the amino acid sequences of each motif. Green letters (motifs IV and VI) indicate that the amino acid differs but its biochemical properties, and therefore its affect on the peptide is similar. However, yellow letters (motif IV) indicate that the differing amino acids do not have similar biochemical properties, and might therefore have an effect of the peptide. Labels above the open boxes are the names assigned to the motifs. The spaces between motifs are arbitrary and not representative of either protein.

Expanding on this observation, the sequences of PFEIF4A and the three human EIF4A isoforms (Appendix 7) were input into the multiple alignment tool of the Clustal program (Thompson *et al.* 1994; Larkin *et al.* 2007) (Appendix 8). This showed that PFEIF4A is most similar to human EIF4A1 and is least similar to human EIF4A3. These results are summarised in Table 4.1.

4.1.2 Substrate Specificity Screen

Helicases have often been shown to have specific preferences for particular substrates, whether favouring DNA over RNA (Lohman *et al.* 1996; de la Cruz *et al.* 1999) or in their resolution of specific nucleic acid topologies (Webb *et al.* 2007; Long *et al.* 2009). Therefore, PFEIF4A purified as detailed in chapter three was used to screen a variety of nucleic acid substrates. An initial screen for helicase activity following the helicase assay as first described by Özsoy *et al.* (2003) was carried out. This was performed with non-phosphorylated PFEIF4A and PFEIF4A phosphorylated at Serine (S) and Threonine (T) residues in response to observations made by Pradhan *et al.* (2008). In short, 20 µL

Protein A	Length A	Protein B	Length (B)	Homology (%)
EIF4A1	406	EIF4A2	407	89
EIF4A1	406	PFEIF4A	398	68
EIF4A2	407	PFEIF4A	398	67
EIF4A1	406	EIF4A3	411	66
EIF4A2	407	EIF4A3	411	66
EIF4A3	411	PFEIF4A	398	63

Table 4.1 **Clustal Results Summary of PFEIF4A and Human EIF4A Isoforms.** The table is arranged in descending percentage homology from top to bottom. The most similar among all four sequences are the human EIF4A isoforms 1 and 2. The second most similar proteins are human EIF4A1 and PFEIF4A, and the third most similar proteins is human EIF4A2 and PFEIF4A. The fourth and fifth most similar proteins are the human isoforms 1 and 2 with isoform 3, and the least similar proteins are PFEIF4A and human EIF4A3.

reactions were performed on 12 substrates with various topologies (Figure 4.2), with 5 μM PFEIF4A and 100 nM DNA substrate (50:1 ratio). Reactions were incubated at 37 °C for 1 hour and stopped with 5 μL of 5 × termination buffer prior to running on SDS PAGE gels (Figure 4.3). In addition, a series of PFEIF4A titrations to these substrates were also carried out at 10, 20, 30, 40, and 50:1 ratios (Figure 4.4).

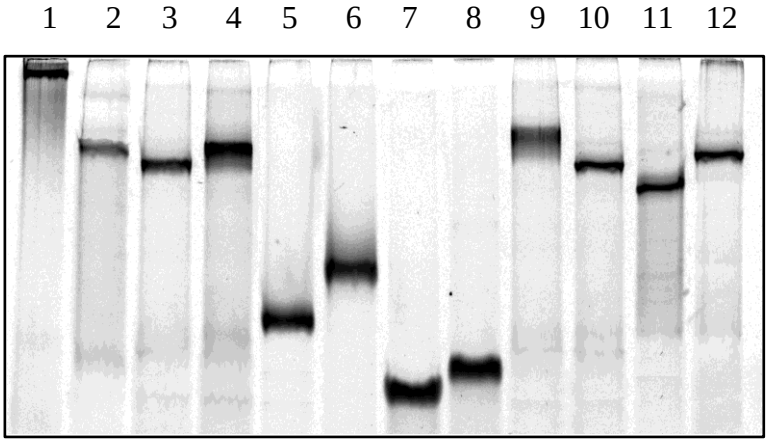


Figure 4.2 **Substrate Specificity Screen Substrates.** 15 % SDS PAGE of substrates 1 to 12 as described in Table 2.3. In lanes 1 to 12 respectively, they are; Holliday Junction, 3 Strand Junction, 3’ Flap, 5’ Flap, 4 bp Bubble, 12 bp Bubble, 36/18 Partial Duplex, 50/19 Partial Duplex, 3 Way Junction, Fork 1, Fork 2, and 80 bp Blunt Duplex.

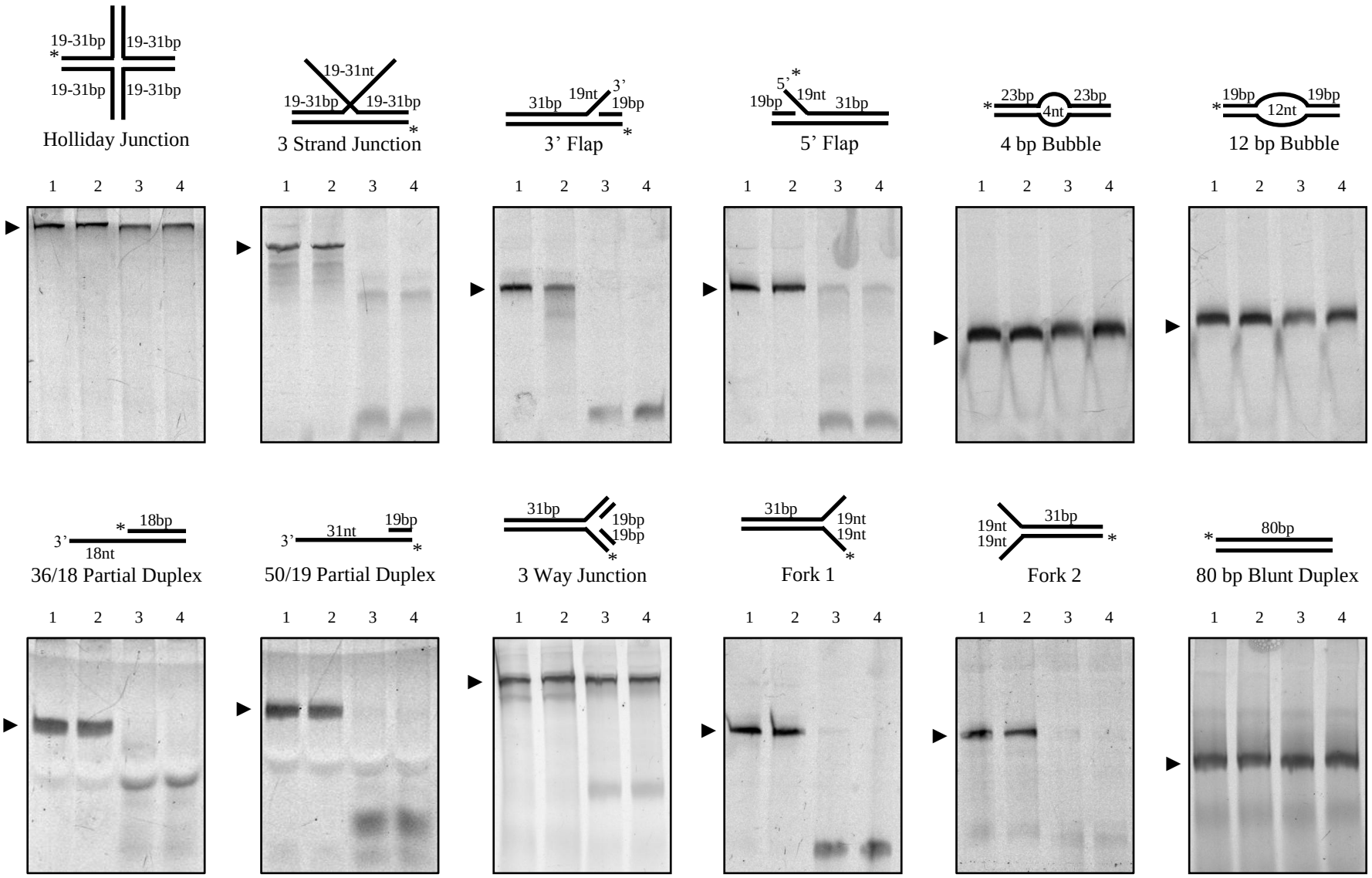

19-31bp
19-31bp
19-31bp
19-31bp
*
Holliday Junction
19-31nt
19-31bp
19-31bp
*
3 Strand Junction
31bp
19nt
3'
19bp
*
3' Flap
5'
*
19bp
19nt
31bp
5' Flap
23bp
23bp
*
4nt
4 bp Bubble
19bp
19bp
*
12nt
12 bp Bubble
1 2 3 4
3'
*
18bp
18nt
36/18 Partial Duplex
3'
31nt
19bp
*
50/19 Partial Duplex
31bp
19bp
19bp
*
3 Way Junction
31bp
19nt
19nt
*
Fork 1
19nt
19nt
31bp
*
Fork 2
*
80bp
80 bp Blunt Duplex

Figure 4.3 **PFEIF4A Substrate Specificity Screen.** 15 % SDS PAGE gels visualised using a Fujifilm FLA-5000 phosphoimager with a blue laser (473 nm) at 550 V and a green filter (532 nm), showing PFEIF4A activity on fluorescein labelled substrates with various topologies as described by Özsoy *et al.* (2003). Above each gel is a representative diagram of the substrate relating to that gel and its associated name as designated by Özsoy *et al.* (2003). The solid black triangle identifies the corresponding band for each of the primary substrates being screened against. Each gel lane is numbered 1 to 4; lanes 1 and 2 contain the room temperature control and the 37°C control respectively, while lane 3 contains the substrate after treating with PFEIF4A, and lane 4 contains the substrate after being treated with phosphorylated PFEIF4A. Repeat experiments show that six of the gels show complete unwinding of the primary substrate after 1 hour at 37°C; 3 Strand Junction, 3' Flap, 36/18 Partial Duplex, 50/19 Partial Duplex, Fork 1, and Fork 2. The 5' Flap also displays unwinding, however this is incomplete after 1 hour. All of the unwound substrates, including the 5' Flap, have a single-stranded overhang.

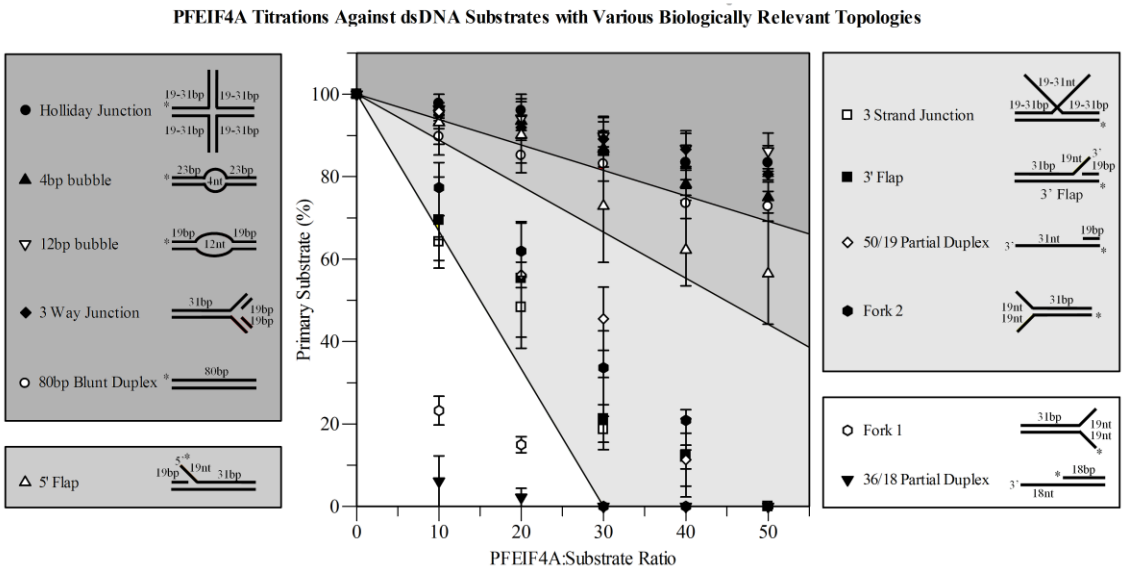


Figure 4.4 **PFEIF4A Substrate Specificity Screen Titrations.** The graph shows the effect of increasing concentration of PFEIF4A on the various substrates used in the substrate specificity screen. The points in the dark grey zone represent those substrates that do not show significant unwinding (<30 %) of the primary substrate at any given concentration. The points in the medium grey zone represent the primary substrates that although show discernible unwinding do not reach completion (30-55 %). Points in the light grey zone represent the primary substrates that reach completion during a 1 hour incubation at a 50:1 PFEIF4A to substrate ratio. Finally, the points in the white zone represent the primary substrates that are unwound to completion at lower PFEIF4A to substrate ratios within the 1hr incubation. The 36/18 Partial Duplex being observed to fully unwind at as low as a 10:1 PFEIF4A to substrate ratio although more reliably at a 30:1 ratio, while the Fork 1 substrate was only ever observed to unwind completely at as low as 30:1 PFEIF4A to substrate ratio. Curiously, Fork 2 is not unwound to the same degree as Fork 1 at the lower ratios. When scrutinised using the OligoAnalyser software, the loose ends of Fork 2 show low levels of complementarity which imparts a more bubble-like rather than forked structure. All zones are drawn arbitrarily based on observable trends; however the given percentages are taken directly from the raw data.

From these experiments it appears that PFEIF4A unwinding activity required a single-stranded overhang and that increasing the concentration of PFEIF4A resulted in increased levels of unwinding on substrates containing single-stranded overhangs; 3 Strand Junction, 3’ Flap, 5’ Flap, 39/18 Partial Duplex, 50/19 Partial Duplex, Fork 1, and Fork 2 substrates. Curiously, the single-stranded regions of the bubble substrates were not sufficient to stimulate PFEIF4A unwinding. Furthermore, under these conditions the 5’ Flap substrate

did not appear fully unwound whereas the 3' Flap substrate did, suggesting a potential 3' polarity bias.

PFEIF4A was phosphorylated at the hydroxyl groups of its serine and threonine residues with Protein Kinase C (PKC). The efficacy of phosphorylation was checked by incorporation of 3.4 fmol ^{32}P-labeled ATP, equating to 1 nanocurie (nCu), and visualised on a SDS PAGE gel (not shown). However, phosphorylation did not appear to affect PFEIF4As specificity to substrates with single-stranded overhangs. As a result, the affects of phosphorylation was not pursued within this study.

4.1.3 PFEIF4A Specific Substrate Screen

After Observing PFEIF4A resolving relatively simple DNA topologies, an additional set of more elaborate substrates inspired by Anand *et al.* (2004) were created (Figure 4.5).

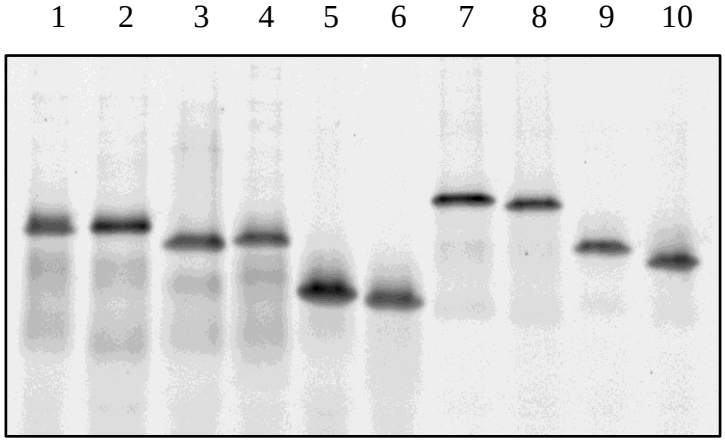


Figure 4.5 **Annealed HPD Substrates.** 15 % SDS PAGE of substrates 13 to 22 as described in Table 2.3. Lanes 1 to 10 represent HPD1 to HPD10 respectively.

These substrates incorporated common features of the unwound Özsoy substrates used in the substrate specificity screen (4.1.2) as well as incorporating hairpin structures. The screen itself was carried out under the same conditions of the substrate specificity screen, described under the helicase assay and run on SDS PAGE gels (Figure 4.6).

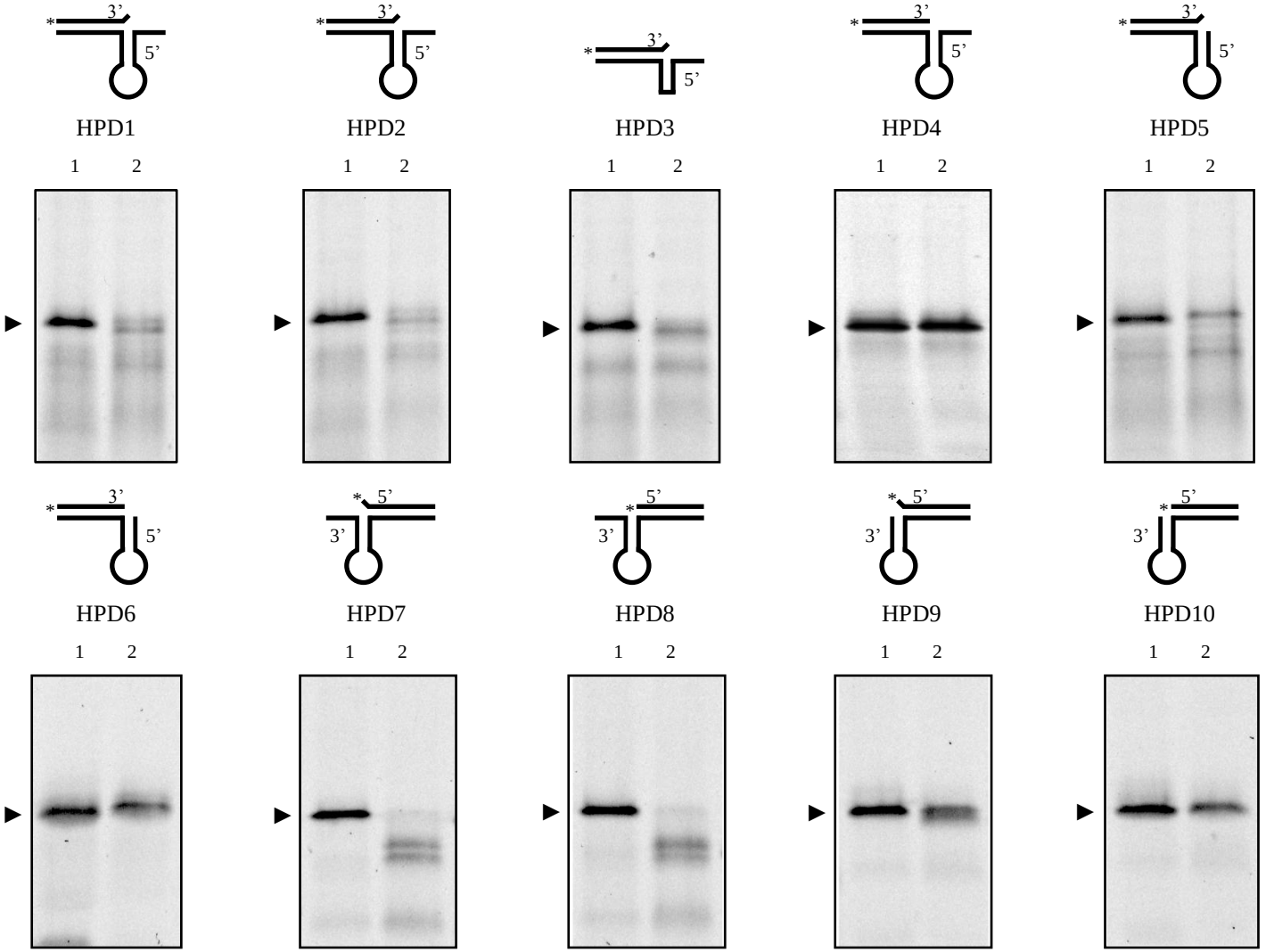

HPD1
HPD2
HPD3
HPD4
HPD5
HPD6
HPD7
HPD8
HPD9
HPD10
1
2
3’
5’
*

Figure 4.6 **PFEIF4A Substrate Specificity Screen.** 15 % SDS PAGE gels visualised using a Fujifilm FLA-5000 phosphoimager with a blue laser (473 nm) at 550 V and a green filter (532 nm), showing PFEIF4A activity on fluorescein labelled substrates with various topologies. Above each gel is a representative diagram of the substrate relating to that gel and its associated name. The solid black triangle identifies the corresponding band for each of the primary substrates being screened against. The lanes on each gel are numbered 1 or 2; lane 1 contains the no helicase control, while lane 2 contains the substrate after being incubated with PFEIF4A. Repeat experiments show that only the gels corresponding to HPD7 and HPD8 show complete unwinding of the primary substrate after 1 hour at 37°C. HPD1, 2, 3, and 5 show partial unwinding, and HPD4, 6, 9, and 10 show no unwinding. The common features of HPD7 and HPD8 is an extending 3' single-stranded overhang, whereas the partially unwound substrates only have a single nucleotide overhang on the 3'. The substrates that showed no unwinding completely lack any form of a 3' single-stranded overhang.

These experiments confirmed that PFEIF4A requires a single-stranded overhang, and blunt ended substrates do not support unwinding. Interestingly, the substrates with 5' single-stranded overhangs or 5' single-nucleotide flaps did not appear to undergo partial unwinding as seen for the 5' Flap substrate described previously (4.1.2). Only substrates with a 3' single-stranded overhang showed complete unwinding by PFEIF4A, while substrates with a 3' single-nucleotide flap were partially unwound, at an average of ~48 %.

4.1.4 Substrate Polarity Screen

The observation that PFEIF4A unwound the 3' Flap more effectively than the 5' Flap, and its inability to resolve the structured substrates with the free 5' single-stranded ends, suggested a potential polarity bias despite previous studies describing PFEIF4A as equally bipolar (Pradhan *et al.* 2007). Therefore a further experiment to investigate polarity was devised.

Two redesigned 50/19 Partial Duplex substrates were used for the polarity experiments. The 3'-5' substrate was identical in sequence to that described by Özsoy *et al.* (2003), however the fluorescein was on the 5' end of the 19mer. This solved the observation of the fluorescein label displacement at the 5' end of oligonucleotides attributing the 3' single-stranded overhang of the substrate (seen for both the 50/19 Partial Duplex and the Fork 2 substrates). This also supports the conclusion that the duplex is being unwound rather than being digested by a contaminating nuclease. The oligonucleotides that made up the 5'-3' substrate were the reverse complement of those that made up the 3'-5' substrate so that the sequence itself could not be accountable for any difference observed. The experiment was performed as a time course with 20 µL 5 minute time points. Each time point reaction was carried out as described for the helicase assay with a 50:1 helicase to substrate ratio (5:0.1 µM), at 37 °C and stopped with 5 µL of 5 × termination buffer (Figure 4.7).

These experiments showed that although both the 3'-5' and the 5'-3' substrates were unwound when incubated with PFEIF4A, only PFEIF4A unwinding activity in the 3'-5' direction was stimulated approximately two-fold in the presence of ATP, from 30.56 ± 1.06×10^{-5} to 62.19 ± 1.65 fmol min^{-1}.

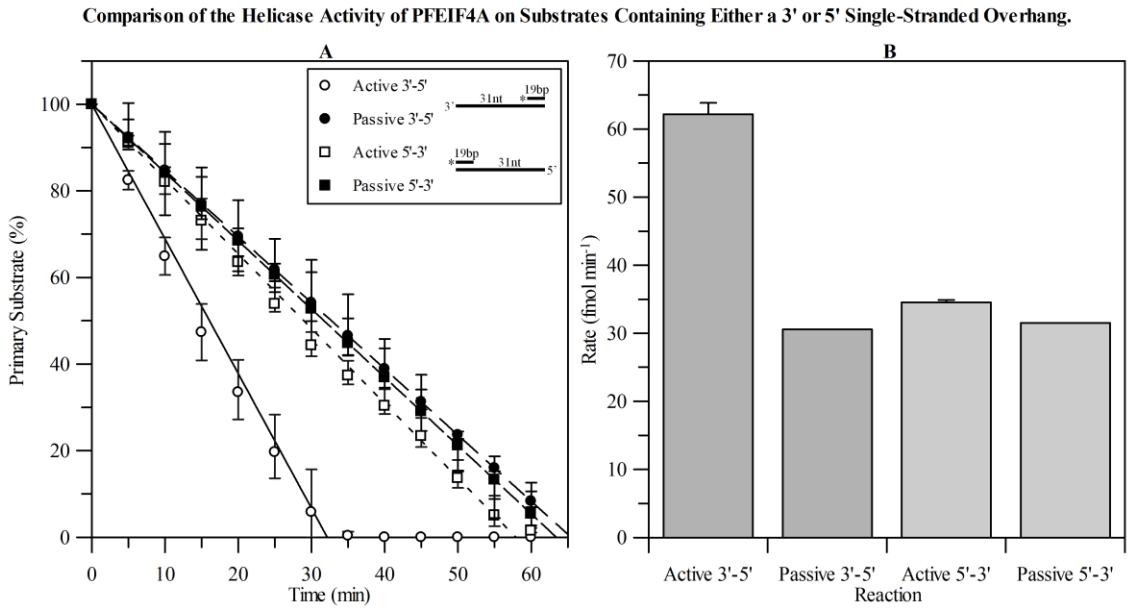


Figure 4.7 **PFEIF4A Substrate Polarity Screen.** Graph A shows the active and passive unwinding activity of PFEIF4A on substrates 23 and 24 as described in Table 2.3. The white circles represent the percentage of the actively unwound 3'-5' primary substrate at the given time points, and the solid line represents their line of best fit. The black circles represent the percentage of the passively unwound 3'-5' primary substrate at the given time points, and the wide-spaced dashed line represents their line of best fit. The white squares represent the percentage of the actively unwound 5'-3' primary substrate at the given time points, and the wide-spaced dotted line represents their line of best fit. Finally, the black squares represent the percentage of the passively unwound 5'-3' primary substrate at the given time points, and the close-spaced dashed line represents their line of best fit. Graph B shows the active and passive unwinding rate (fmol min^{-1}) of PFEIF4A for the 3'-5' 50/19 Partial Duplex (dark grey) and the 5'-3' 50/19 Partial Duplex (light grey).

4.1.5 Single-Stranded Overhang Requirements

As the single-stranded regions within the bubble substrates did not appear sufficient to support PFEIF4A unwinding activity, and the 3' single-nucleotide structured substrates were not unwound to completion, the absolute requirement for a single-stranded overhang needed to be explored. Given PFEIF4As preferential unwinding activity in the 3'-5' direction, a series of oligonucleotides were designed to create multiple partial duplex substrates with decreasing single-stranded overhangs based on truncated forms of the 3'-5' substrate used in the substrate polarity screen (Figure 4.8).

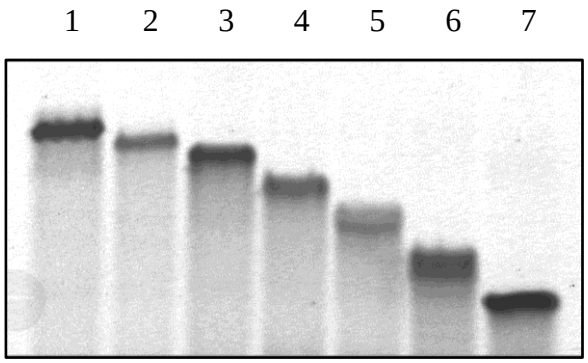


Figure 4.8 **Truncated Partial Duplex Substrates.** 15 % SDS PAGE gel of substrates 23 and 25 to 30 as described in Table 2.3. In lanes 1 to 7 respectively, they are; 50/19 Partial Duplex, 45/19 Partial Duplex, 40/19 Partial Duplex, 35/19 Partial Duplex, 30/19 Partial Duplex, 25/19 Partial Duplex, and 20/19 Partial Duplex.

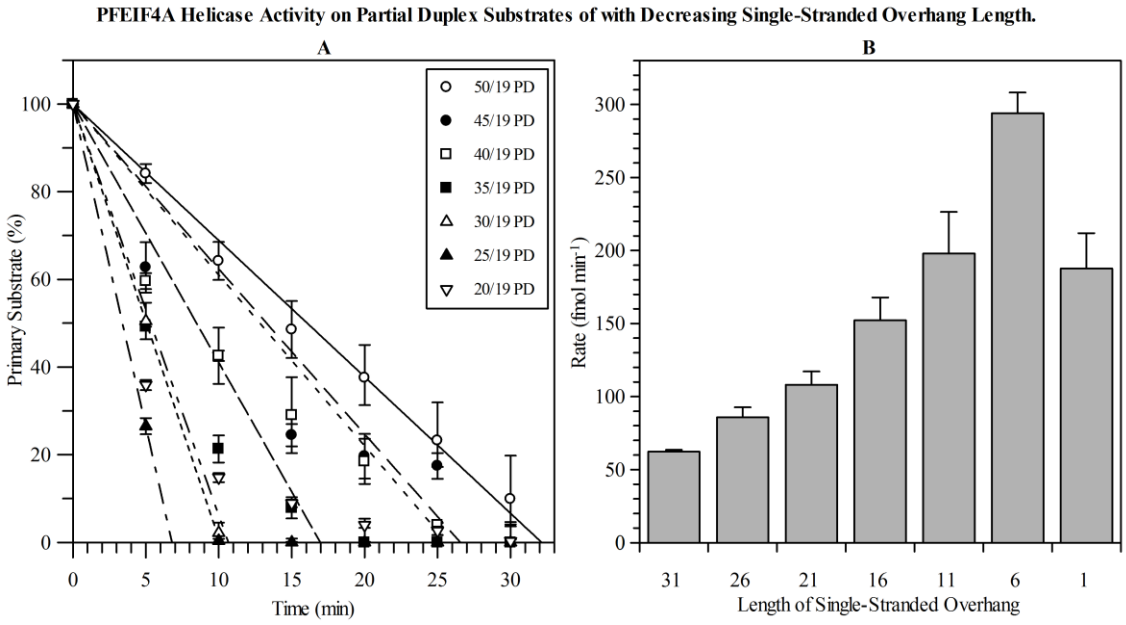


Figure 4.9 **Single-Stranded Overhang Requirements of PFEIF4A.** Graph A shows the initial unwinding rates of PFEIF4A on the various Partial Duplex substrates. The percentage of the primary substrate for the 50, 45, 40, 35, 30, 25, and 20/19 Partial Duplex substrates at the given time points are represented by the white circles, black circles, white squares, black squares, white triangles, black triangle, and the inverted white triangles respectively. The initial rates of the reactions for each substrate are represented by the solid (50/19), the wide-spaced dashed (45/19), the wide-spaced dotted (40/19), the close-spaced dashed (35/19), the close-spaced dotted (30/19), the dot-dashed (25/19), and the dot-dot-dashed lines (20/19). The comparative unwinding rates (fmol min^{-1}) of PFEIF4A for each of the Partial Duplex substrates are shown in Graph B.

Time course experiments were carried out as described for the polarity experiments, with 20 μL 5 minute time points (Figure 4.9). These showed that PFEIF4A unwinding experiments on substrates with shorter single-stranded overhangs reached completion quicker than those with longer single-stranded overhangs. The exception was the 20/19 Partial Duplex substrate with a 1 nucleotide (nt) overhang (187.59 ± 24.30 fmol min^{-1}), which took longer than both the 30/19 (198.02 ± 28.42 fmol min^{-1}) and 25/19 (294.01 ± 14.21 fmol min^{-1}) Partial Duplex substrates with 11 and 6 nt overhangs respectively.

4.1.6 Ribonucleic Acid (RNA) Unwinding Activity

EIF4A is a well researched RNA helicase with plant and animal orthologues, and was first purified in the 1970s (Rogers *et al.* 2002). Despite EIF4As ability to unwind RNA:DNA and DNA:RNA hybrid substrates, it does not seem capable of unwinding dsDNA (Rogers *et al.* 2001). Given that all the substrates used in this study so far had been dsDNA, new substrates, based on the Fork 1 substrate, were designed to reveal any difference between PFEIF4A activity against DNA:RNA and RNA:DNA heteroduplexes, and dsRNA (Figure 4.10).

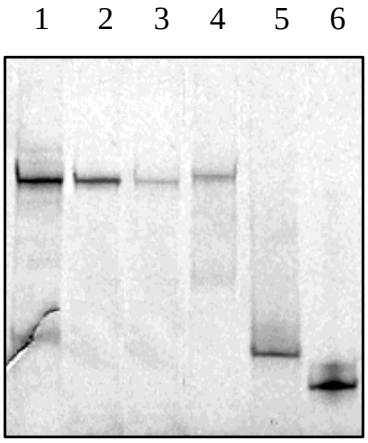


Figure 4.10 **DNA and RNA Fork 1 Substrates.** 15 % SDS PAGE gel of substrates 31 to 33 as described in Table 2.3. Lanes 1 to 4 show the dsDNA, the DNA:RNA and RNA:DNA hybrids, and the dsRNA Fork substrates respectively. Lane 5 and 6 show the 5' fluorescein labelled oligonucleotides 4 and 31 respectively (Appendix 1).

Time course experiments were performed to examine PFEIF4A activity against the DNA, RNA and hybrid Fork 1 substrates. Due to the observed thermal instability of the RNA:DNA hybrid, a single room-temperature time point control at zero minutes and 37 °C controls at the subsequent time points were taken. In addition, due to concerns about the hybrid substrates stability, the reactions were stopped with 5 µL of 5 × modified

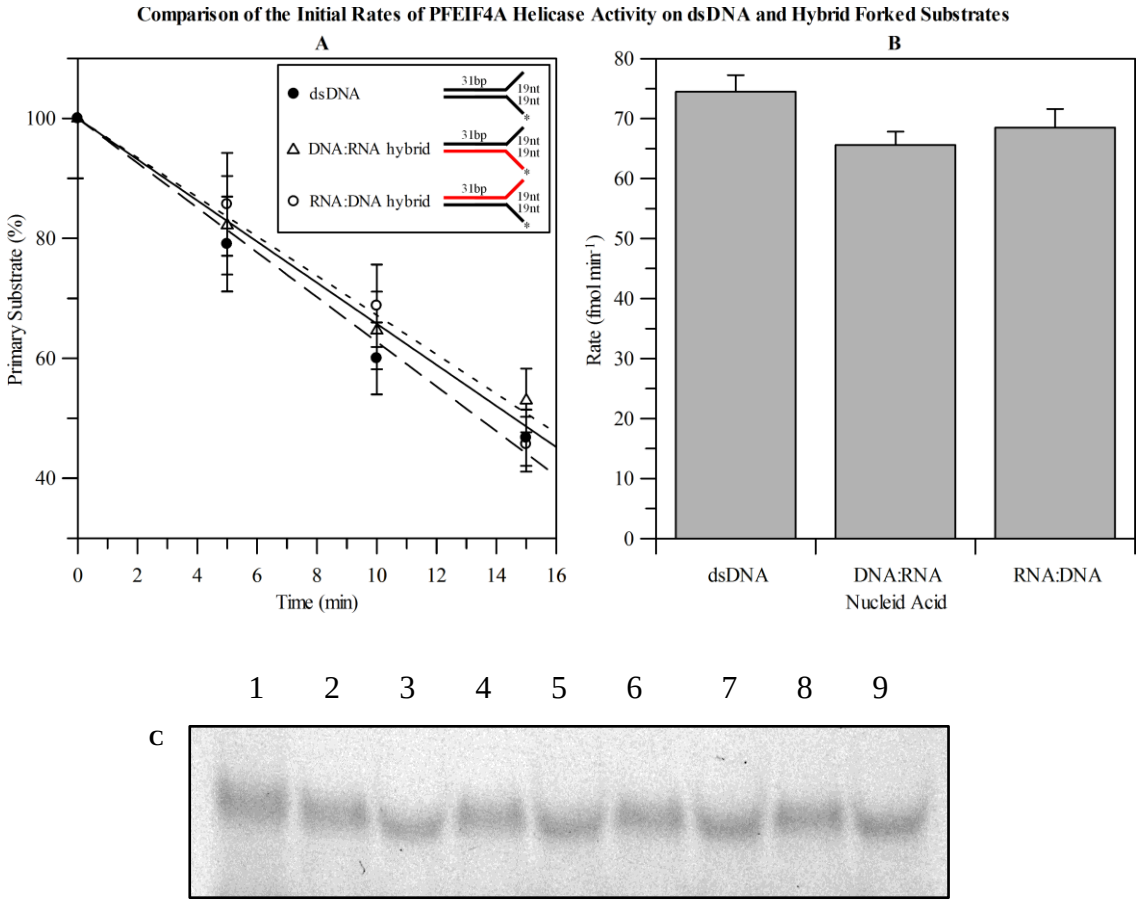


Figure 4.11 **PFEIF4A Ribonucleic Acid (RNA) Unwinding Activity.** Graph A shows the unwinding activity of PFEIF4A on DNA, RNA, and DNA:RNA and RNA:DNA hybrids. The white circles represent the percentage of the primary substrate for the RNA:DNA hybrid substrate at the given time points, and the solid line represents their line of best fit. The black circles represent the percentage of the primary substrate for the dsDNA substrate at the given time points, and the wide-spaced dashed line represents their line of best fit. Graph B shows the comparative unwinding rates (fmol min^{-1}) of PFEIF4A on the RNA:DNA hybrid and dsDNA forked substrates. The 6 % native PAGE gel (C) is an example result for the dsRNA substrate. Lane 1 shows the 0 minute control, lanes 2, 4, 6, and 8 show the 5, 10, 15, and 20 minute control samples respectively, while lanes 3, 5, 7, and 9 show the samples incubated with PFEIF4A at 5, 10, 15, and 20 minutes respectively.

termination buffer, which did not contain SDS. Furthermore, the reactions were run on native PAGE gels at 100 V for 1 hour at 4 °C (Figure 4.11).

These experiments show that PFEIF4A unwinds DNA:RNA and RNA:DNA hybrids, and dsDNA substrates of equal topology to the same degree; 65.58 ± 2.25, 68.49 ± 3.12, and 74.49 ± 2.76 fmol min^{-1} respectively. However, it appeared that the dsRNA substrate was not unwound in the presence of PFEIF4A.

4.1.7 Cofactor Requirement

Interest in the general requirements of PFEIF4A for full activity led to experiments being designed to elucidate PFEIF4As divalent metal cofactor requirement. In these experiments, rather than using the assay buffer as described under the helicase assay, buffers where the $MgCl_2$ was substituted with alternative biologically relevant metal salts were used. Under these modified conditions, a series of 50:1 PFEIF4A to 36/18 Partial Duplex ratio (5:0.1 μM) time courses with 20 μL 5 minute time points were performed (Figure 4.12).

These showed that the most effective cofactor for PFEIF4A helicase activity was Mg^{2+} cations (175.48 ± 2.04 fmol min^{-1}), and although unwinding activity was observed in the presence of Mn^{2+} cations it was notably down-regulated (34.39 ± 1.57 fmol min^{-1}). Neither Zn^{2+} or Ca^{2+} cations supported unwinding, and the EDTA control show total inhibition of helicase activity.

4.1.8 Optimum Working Temperature

Additional experiments were also carried out to determine PFEIF4A's optimum working temperature range. Reagents were pre-incubated at various temperatures (22, 27, 30, 34, 37, 40, and 45 °C) prior to carrying out the time course reactions.

20 μL 5 minute time points were set-up under the conditions described under the helicase assay with a 50:1 PFEIF4A to 36/18 Partial Duplex ratio (5:0.1 μM) (Figure 4.13).

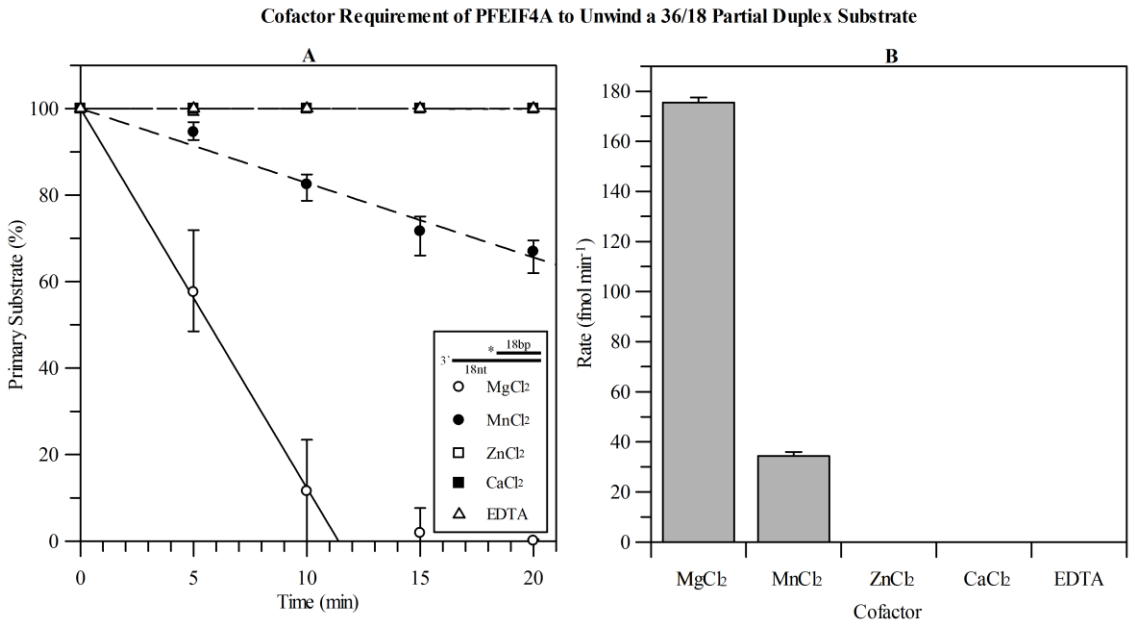


Figure 4.12 **PFEIF4A Cofactor Requirements.** Graph A shows the unwinding activity of PFEIF4A on substrate 7 as described in Table 2.3. The white circles represent the percentage of the primary substrate at the given time points in the presence of Mg^{2+}, and the solid line represents their line of best fit. The black circles represent the percentage of the primary substrate at the given time points in the presence of Mn^{2+}, and the wide-spaced dashed line represents their line of best fit. The white squares represent the percentage of the primary substrate at the given time points in the presence of Zn^{2+}, and the wide-spaced dotted line represents their line of best fit. The black squares represent the percentage of the primary substrate at the given time points in the presence of $CaCl_2$, and the close-spaced dashed line represents their line of best fit. The white triangles represent the percentage of the primary substrate at the given time points in the presence of EDTA, and the close-spaced dotted line represents their line of best fit. The comparative unwinding rates (fmol min^{-1}) of PFEIF4A in the presence of $MgCl_2$ and $MnCl_2$ are shown in Graph B.

The results show that PFEIF4A has a broad active temperature range between 27-37 °C where the extent of unwinding was similar, although 37 °C appears to be the optimum (175.42 ± 2.01 fmol min^{-1}). Interesting, despite the flexibility in the lower range temperatures, activity was markedly lower at 40 °C and above (50.64 ± 2.83 fmol min^{-1}). Furthermore, experiments carried at 37 °C using PFEIF4A pre-incubated at 40 °C reveals that the loss in activity was not recoverable.

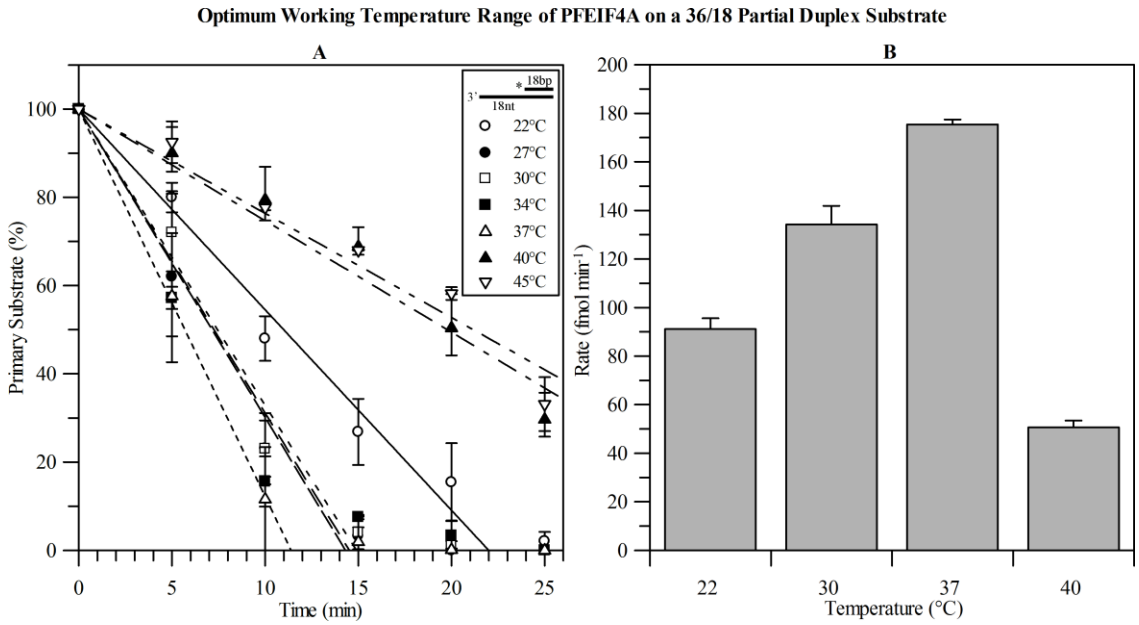


Figure 4.13 **PFEIF4A Optimum Working Temperatures.** Graph A shows the degree of unwinding at various temperatures by PFEIF4A on substrate 7 as described in Table 2.3. The percentage of the primary substrate at the given time points for the reaction temperatures 22, 27, 30, 34, 37, 40, and 45 °C, are represented by the white circles, black circles, white squares, black squares, white triangles, black triangles, and inverted white triangles respectively. The initial rates of the reactions at each temperature are represented by the solid (22 °C), the wide-spaced dashed (27 °C), the wide-spaced dotted (30 °C), the close-spaced dashed (34 °C), the close-spaced dotted (37 °C), the dot-dashed (40 °C), and the dot-dot-dashed lines (45 °C). Graph B illustrates in the increased rate of unwinding (fmol min^{-1}) from 22 to 37 °C followed by the dramatic decrease at 40 °C.

4.1.9 Nucleotide Triphosphate (NTP) Usage

The 3'-5' 50/19 Partial duplex described in Figure 4.7 was used to explore PFEIF4As NTP usage. Time course reactions with 20 µL 5 minute time points were set-up under the conditions described for the helicase with a 50:1 ratio (5:0.1 µM) of PFEIF4A to 50/19 Partial Duplex (3'-5'). These experiments were supplemented with or without the presence of ATP, TTP, GTP, CTP, UTP, dATP, dTTP, dGTP, dCTP, di (ADP) and mono (AMP) adenosine phosphate, and the ATP analogues adenosine 5'-O-(3-thio)triphosphate (ATPγS) and adenosine 5'-(β,γ-imido)triphosphate (AMP-PnP) (Figure 4.14).

These experiments showed in addition to ATP (63.55 ± 2.67 fmol min^{-1}), dCTP (53.04 ± 1.24 fmol min^{-1}) and dGTP (37.89 ± 0.89 fmol min^{-1}) showed a marked increase in

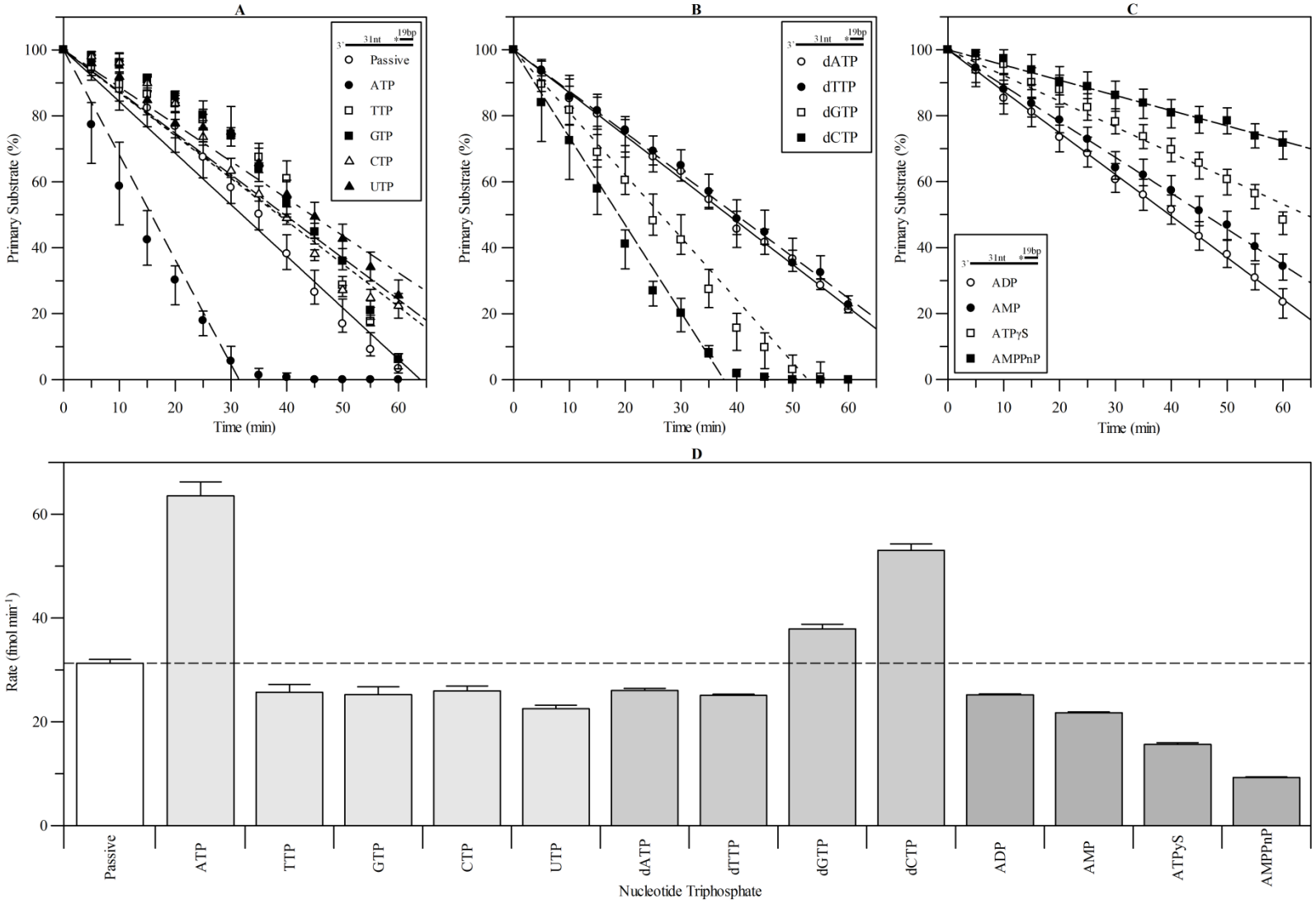

The Effects of Various Nucleotides on PFEIF4A Helicase Activity on a 50/19 Partial Duplex Substrate.
A
B
C
D
Primary Substrate (%)
Time (min)
3' 31nt 19bp 5'
Passive
ATP
TTP
GTP
CTP
UTP
dATP
dTTP
dGTP
dCTP
ADP
AMP
ATPγS
AMPPnP
Rate (fmol min-1)
Nucleotide Triphosphate

Figure 4.14 **PFEIF4A Nucleotide Triphosphate (NTP) Usage.** Graphs showing PFEIF4A unwinding 50/19 Partial Duplex (3'-5') (substrate 23 in Table 2.3) in the presence of various NTPs, dNTPs, and ATP analogues. Graph A shows the percentage of the primary substrate at the given time points under passive and active (ATP) unwinding conditions, represented by the white circles with the solid line of best fit and the black circles with the wide-spaced dashed line of best fit respectively. On the same graph, the percentage of the primary substrate at the given time points for unwinding in the presence of TTP, GTP, CTP, and UTP, are represented by the white squares, black squares, white triangles, and black triangles respectively. Their lines of best fit are represented by the wide-spaced dots (TTP), close-spaced dashes (GTP), close-spaced dots (CTP), and dot-dashes (UTP). Graph B shows the percentage of the primary substrate at the given time points in the presence of dATP (white circles), dTTP (black circles), dGTP (white squares), and dCTP (black squares). Their lines of best fit are represented by the solid line, the wide-spaced dashes, the wide-spaced dots, and the close-spaced dashes respectively. Graph C shows the percentage of the primary substrate at the given time points in the presence of ADP (white circles), AMP (black circles), ATPγS (white squares), and AMP-PnP (black squares). The solid, wide-spaced dashes, wide-spaced dots, and the close-spaced dashes illustrate their lines of best fit respectively. The unwinding rates (fmol min^{-1}) of PFEIF4A in the presence of each NTP, dNTP, and ATP analogues are represented graph D. PFEIF4As passive rate is shown in white as the left-most column and the dashed line represents its rate for easy comparison. All other columns are shaded according to their respective graphs; light grey columns are the rates taken from graph A, medium grey columns are the rates taken from graph B, and the dark grey columns are the rates taken from graph C.

PFEIF4As unwinding activity when compared to its passive activity (31.28 ± 0.75 fmol min^{-1}). Furthermore, a marked inhibition of PFEIF4As passive activity was observed when in the presence of ATPγS (15.61 ± 0.36 fmol min^{-1}) and AMP-PnP (9.24 ± 0.14 fmol min^{-1}).

4.1.10 Electrophoretic Mobility Shift Assay (EMSA)

Based on the observations made in the previous section (4.1.9) that PFEIF4A passive activity was inhibited in the presence of AMP-PnP, gel shifts exploring PFEIF4As' binding affinity for ssDNA and dsDNA, and for substrates with one single-stranded overhang and two single-stranded overhangs were performed. 20 µL reactions were set-up under the conditions described for the helicase assay, differing only in that ATP was substituted for AMP-PnP. Based on low interactions observed previously at ratios less than 50:1, reactions for each substrate were performed at 50, 100, 200, 300, and 400:1 PFEIF4A to DNA concentrations (Figure 4.15).

This showed that PFEIF4A has a relatively high affinity for DNA substrates with one or more free single-stranded ends (K_d = 4.0313 ± 0.9076 µM) compared to a blunt ended dsDNA substrate (K_d = 45.7427 ± 11.6439 µM).

Helicase assays have shown that PFEIF4A unwinds DNA, RNA, and hybrid substrates to the same extent (4.1.6). Additional gel shift experiments were carried out in order to scrutinise any binding preference to DNA or RNA. Experiments were performed as described above, using single-stranded and double-stranded DNA and RNA substrates.

These experiments revealed that PFEIF4As' binding affinity for RNA is similar to that of it binding affinity for DNA. With a K_d of 5.4272 ± 0.4457 and 4.6803 ± 0.7250 µM for ssRNA and dsRNA respectively, compared to 8.974 ± 0.8325 and 5.7759 ± 0.9235 µM for ssDNA and dsDNA respectively (Figure 4.16).

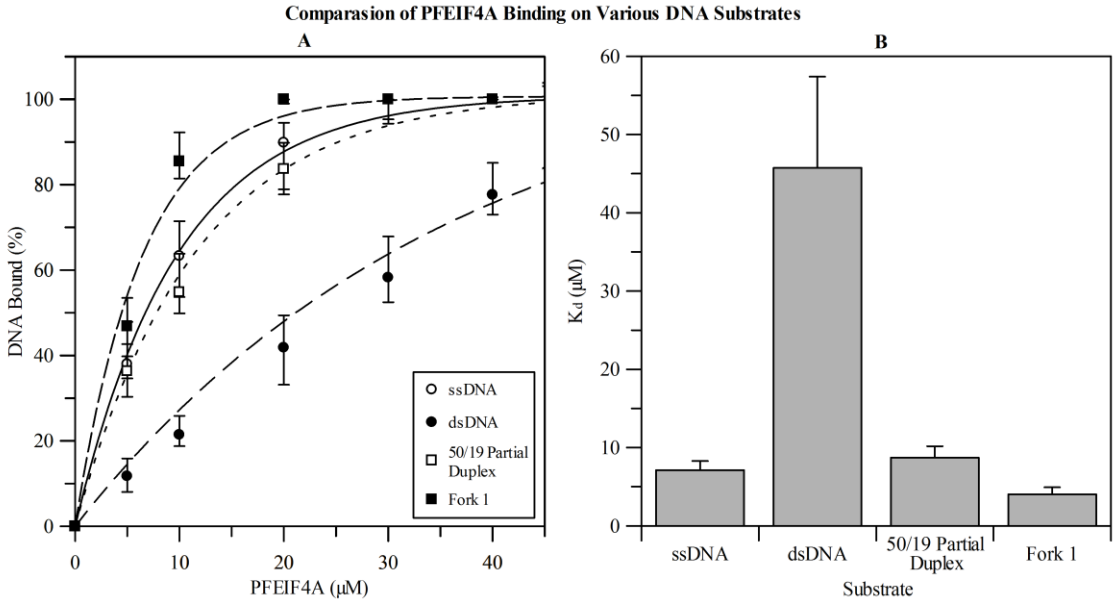


Figure 4.15 **Comparative Analysis of PFEIF4A Binding.** Graph A shows the percentage of DNA bound by PFEIF4A at the given concentrations along the *x*-axis (μM). The white circles represent the percentage of ssDNA (oligonucleotide 11, Appendix 1) bound, and the solid line represents their line of best fit. The black circles represent the percentage of dsDNA (substrate 12, Table 2.3) bound, and the wide-spaced dashes represent their line of best fit. The white squares represent the bound percentage of the 50/19 Partial Duplex substrate (substrate 23, Table 2.3), and the wide-spaced dots represent their line of best fit. The black squares represent the bound percentage of the Fork 1 substrate (substrate 10, Table 2.3), and the close-spaced dashes represent their line of best fit. Data is fit to a single site ligand binding curve. Graph B shows the associated K_d measurements (μM). This reveals a relatively higher affinity of PFEIF4A to DNA substrates with ssDNA regions; 7.1142 ± 1.1836, 8.7037 ± 1.4786, and 4.0313 ± 0.9076 for ssDNA, the 50/19 Partial Duplex, and the Fork1 substrate respectively, versus blunt ended DNA substrates (45.7427 ± 11.6439).

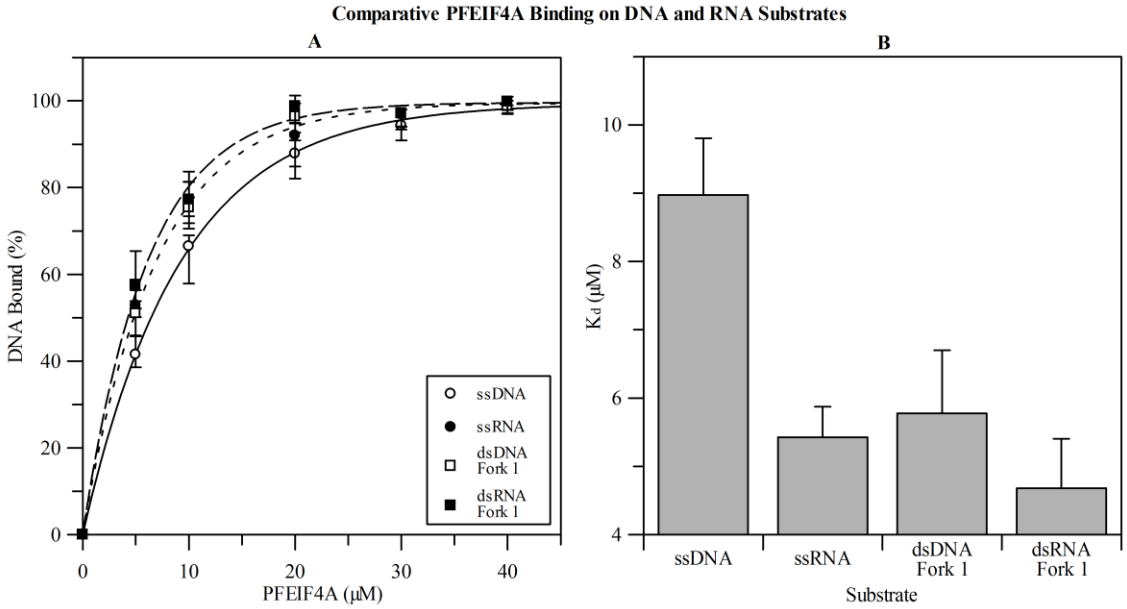


Figure 4.16 **Comparative Analysis of PFEIF4A Binding to DNA and RNA Substrates.** Graph A shows the percentage of the substrates bound by PFEIF4A at the given concentrations along the *x*-axis (μM). The white circles represent the percentage of ssDNA **(**oligonucleotide 4**,** Appendix 1) bound, and the solid line represents their line of best fit. The black circles represent the percentage of ssRNA (oligonucleotide 25**,** Appendix 1**)** bound, and the wide-spaced dashes represent their line of best fit. The white squares represent the bound percentage of the dsDNA Fork 1 substrate (substrate 10, Table 2.3), and the wide-spaced dots represent their line of best fit. The black squares represent the bound percentage of the dsRNA Fork 1 substrate (substrate 27, Table 2.3), and the close-spaced dashes represent their line of best fit. Data is fit to a single site ligand binding curve. Graph B shows the associated K_d measurements (μM). This reveals that PFEIF4A has the lowest affinity to the ssDNA substrates, with a K_d of 8.974 ± 0.8325, while its affinity for ssRNA and both the dsDNA and dsRNA Fork 1 substrates are all similar; 5.4272 ± 0.4457, 5.7759 ± 0.9235, and 4.6803 ± 0.7250 respectively.

4.1.11 ATPase Activity

PFEIF4A's ATPase activity in the presence of single-stranded and double-stranded DNA, as well as the Fork 1 substrate was measured with a series of reverse time courses at 50:1 molar ratios of PFEIF4A (50 μM) to substrate (1 μM). Each time course was performed as described under the malachite green ATPase assay. The resulting data was corrected for background ATP degradation, calculated from successive time points using the same parameters as the Fork 1 time course, only without any helicase present (Figure 4.17).

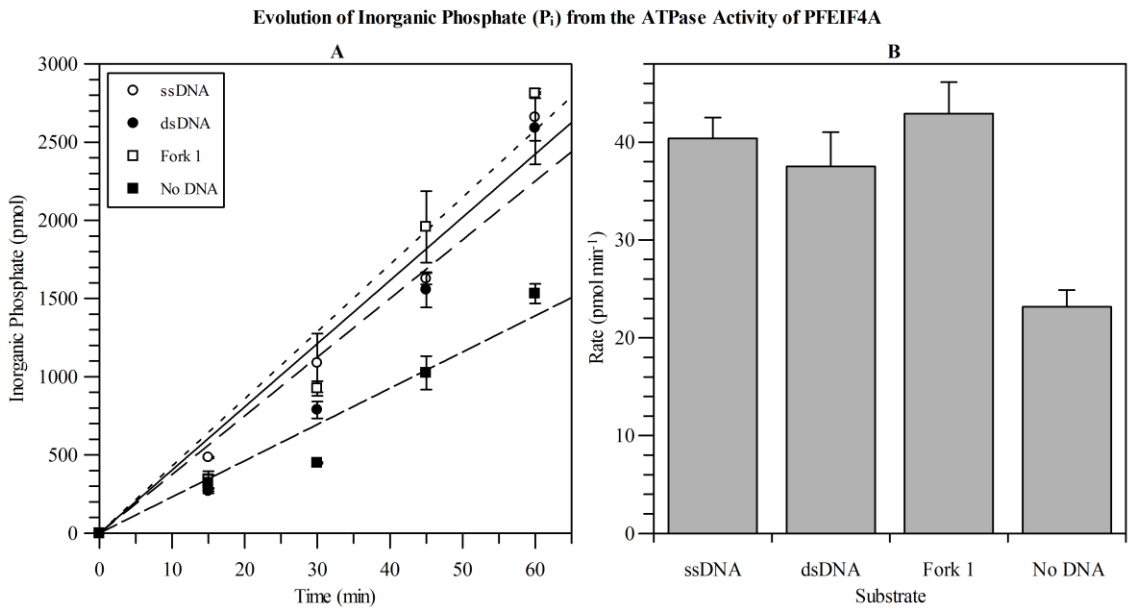


Figure 4.17 **PFEIF4A ATPase Activity measured using the Malachite Green Assay.** Graph A shows the amount of evolved inorganic phosphate (P_i) from the hydrolysis of ATP by PFEIF4As ATPase activity. The white circles represent the amount of P_i evolved in the presence of ssDNA (oligonucleotide 11, Appendix 1), and the solid line represents their line of best fit. The black circles represent the amount of P_i evolved in the presence of dsDNA (substrate 12, Table 2.3), and the wide-spaced dashes represent their line of best fit. The white squares represent the amount of P_i evolved in the presence of the Fork 1 substrate (substrate 10, Table 2.3), and the wide-spaced dots represent their line of best fit. The black squares represent the amount of P_i evolved without any DNA present, and the close-spaced dashes represent their line of best fit. Graph B shows the rate of inorganic phosphate evolved (pmol min^{-1}), revealing an idle ATPase rate (without DNA) of 23.16±1.74 pmol min^{-1}. The turnover rate was enhanced by the addition of DNA, with a comparative increase to 40.38±2.12 pmol min^{-1} in the presence of ssDNA, 37.50±3.51 pmol min^{-1} for dsDNA, and 42.92±3.21 pmol min^{-1} for the Fork 1 substrate.

These experiments show that PFEIF4A has a background ATP turnover rate of 23.16 ± 1.74 pmol min^{-1}. Despite previous observations suggesting that blunt end dsDNA was not unwound, the comparative levels of evolved P_i from ATPase activity in the presence of 80 bp Blunt Duplex, as well as ssDNA and the Fork 1 substrate, doubles to 42.92 ± 3.21 pmol min^{-1}.

The luciferase ATPase assay was carried out alongside the malachite green ATPase assay. However, due to its higher level of sensitivity it could be performed under the standard

helicase conditions of 5 µM PFEIF4A, 0.1 µM DNA substrate, and 1 mM ATP (Figure 4.18).

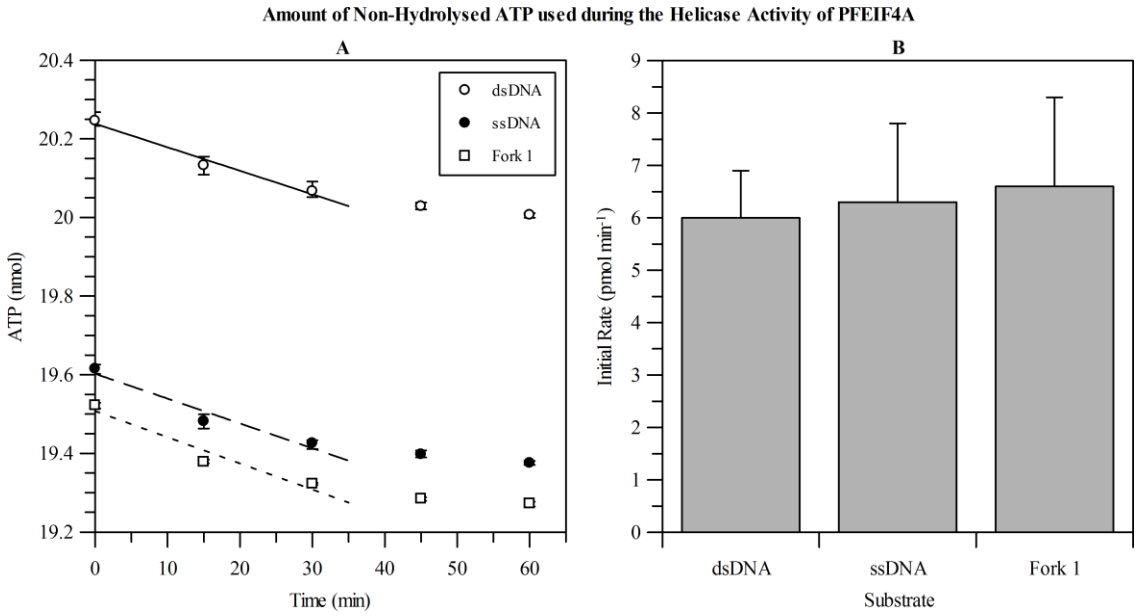


Figure 4.18 **PFEIF4A ATPase Activity measured using the Luciferase Assay.** Graph A shows the amount of unused ATP during PFEIF4A ATPase activity. The white circles represent the concentration of ATP (nmol) at the given time points in the presence of dsDNA (substrate 12, Table 2.3), and the solid line represents their line of best fit for minutes 0 to 35. The black circles represent the concentration of ATP at the given time points in the presence of ssDNA (oligonucleotide 11, Appendix 1), and the wide-spaced dashes represent their line of best fit for minutes 0 to 35. Finally, the white squares represent the concentration of ATP at the given time points in the presence of the Fork 1 substrate (substrate 10, Table 2.3), and the wide-spaced dots represent their line of best fit for minutes 0 to 35. Graph B shows the initial rates of ATP hydrolysis (pmol min^{-1}); 6.00 ± 0.90, 6.30 ± 1.50, and 6.60 ± 1.70 pmol min^{-1} in the presence of dsDNA, ssDNA, and the Fork 1 substrate respectively.

Despite giving slightly higher ATP hydrolysis rates (6.00 ± 0.90 pmol min^{-1}), when taking into account initial ATP concentrations for each assay, it also showed that PFEIF4A ATPase activity was not affected by DNA substrate topology. However, it did not report any background ATP turnover.

4.1.12 Triplex Displacement Assay

Based on the observation that PFEIF4A ATPase activity is stimulated in the presence of the 80 bp Blunt Duplex substrate, despite appearing not to unwind blunt end DNA duplexes, helicase assays were performed on blunt ended dsDNA substrates with either a 5' or 3' triplex. The triplex forming oligonucleotide (TFO) was ^{32}P-radiolabelled by Mervyn Middleton (University of Portsmouth, Hampshire, UK) and the triplexes were formed as described by Firman *et al*. (2000) using the blunt end dsDNA substrates 34 and 35, containing a 5' and 3' triplex binding site respectively (Table 2.3).

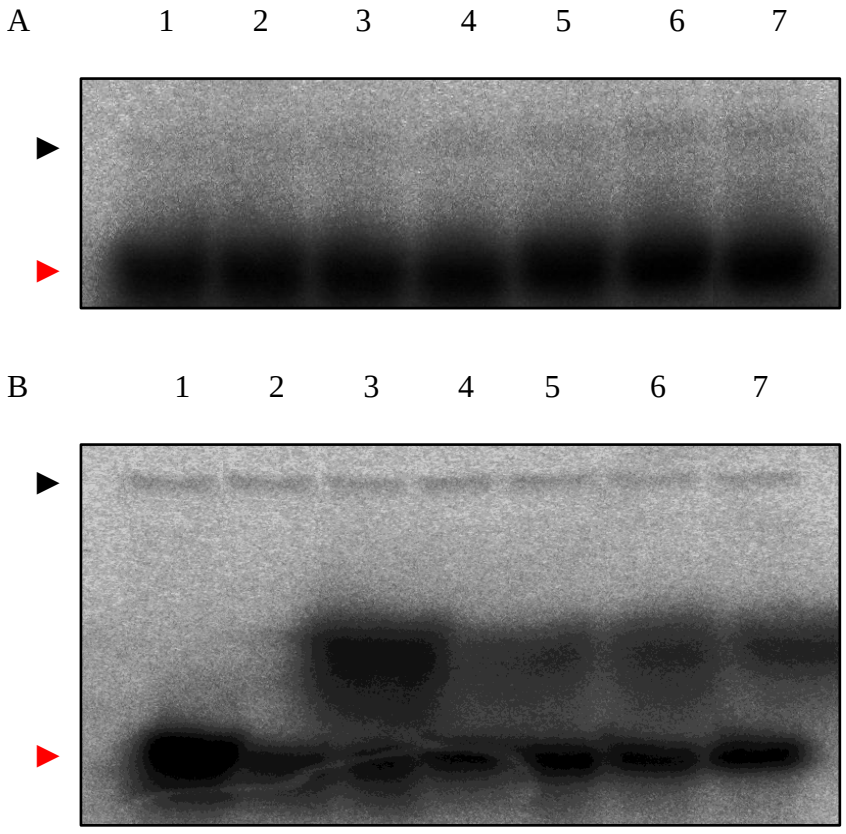


Figure 4.19 **Triplex Displacement from a Blunt-End dsDNA Substrate by PFEIF4A.** 6% native PAGE gels showing dsDNA substrates (substrates 34 and 35, Table 2.3) with either a 3' bound triplex (A) or a 5' bound triplex (B). The lanes on both gels are labelled 1 to 7. Lane 1 contains the 0 minute control sample, while lanes 2 to 7 contain the 10, 20, 30, 40, 50, and 60 minute time point samples. For both gels, the triplex substrate is indicated by the black arrow, while the free ^{32}P-TFO is indicated by the red arrow. Triplex displacement is not observed on either the 3' or 5' triplex substrate.

1 hour time course experiments with 10 minute time points were carried out at a 50:1 PFEIF4A to substrate molar ratio (500:10 nM). The reactions at each time point were stopped with 5 × modified termination buffer, did not contain SDS, and the samples were run on native PAGE gels (Figure 4.19).

The results show that the TFO was not displaced from either of the two duplex DNA substrates when incubated with PFEIF4A.

4.1.13 Developing a High-Throughput Helicase Assay

The gel based nature of the helicase assay used thus far made it a relatively drawn out process. Now that sufficient biochemical characterisation had been performed, an alternative high-throughput protocol was sought after for future work.

DNA binding dyes can be detected and recorded in real-time using a fluorescence microplate reader, effectively eliminating the error-prone aspect of extracting the gel data using densitometry. Several DNA reporter compounds were researched, leading to DNA binding dyes; Hoechst (Latt *et al.* 1975), PicoGreen (Ahn *et al.* 1996), and BOXTO (Eriksson *et al.* 2003; Karlsson *et al.* 2003) being investigated. In preliminary experiments, 20 µL reactions containing the dyes with various duplex DNA substrates were assessed. These used the same conditions and buffer parameters of the gel based helicase assay (excluding the presence of a helicase) and were carried out and visualised on native PAGE gels using a Fujifilm FLA-5000 phosphoimager with a blue laser (473 nm) at 550 V and a green filter (532 nm). BOXTO gave the best measurable results, with a distinguishable difference between the boiled samples (in ssDNA from) and the intact (dsDNA) substrates at a 5:1 BOXTO to DNA molar ratio. Furthermore, this relationship could be reproduced on the fluorescent microplate reader (Figure 4.20).

When a helicase assay was attempted, the presence of the protein affected the fluorescence of BOXTO, making the helicase unwinding sample and the non-helicase control sample difficult to compare during data interpretation. This was overcome based on the observation that Mg^{2+} was required for unwinding to take place (4.1.7), by making up the

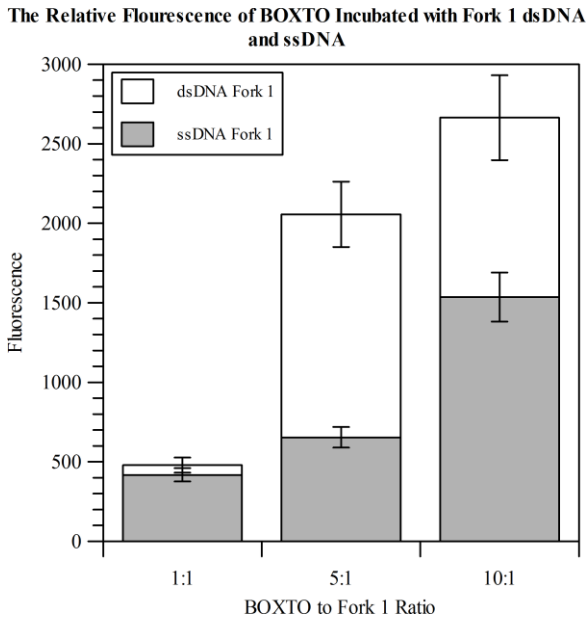


Figure 4.20 **BOXTO Incubated with the Fork 1 Substrate.** The graph shows the relative fluorescence of the Fork 1 substrate (dsDNA) shown in white and its boiled counterpart (ssDNA) shown in grey. The 5:1 BOXTO to Fork 1 substrate incubation shows the greatest discrimination between single- and double-stranded DNA, while the 1:1 shows very little, and the 10:1 shows similar discrimination but also shows an increase in background fluorescence associated with ssDNA.

control sample with the helicase present but substituting the Mg^{2+} for EDTA in the assay buffer (Figure 4.21). The final protocol is outlined in chapter two.

4.2 PFDH60

Some characterisation of PFDH60 had been undertaken (Pradhan *et al.* 2005; Pradhan *et al.* 2006; Pradhan *et al.* 2007; Pradhan *et al.* 2008), but as mentioned previously, these studies used a truncated form of the protein (58.8 kDa). *In silico* techniques revealed shared identity with the DEAD-box containing helicases. Optimal activity was observed in the presence of 1 mM ATP, 30 mM KCl, and 1 mM $MgCl_2$ (Pradhan *et al.* 2005; Pradhan *et al.* 2007). However it has comparable activity in the presence Mn^{2+} in place of Mg^{2+} (Pradhan *et al.* 2007). It was active across a broad pH range (Pradhan *et al.* 2005; Pradhan *et al.* 2007), it was expressed in the Schizont life stage of the *Plasmodium falciparum* parasite (Pradhan *et al.* 2005; Pradhan *et al.* 2007), it was bipolar, unwinds duplex DNA

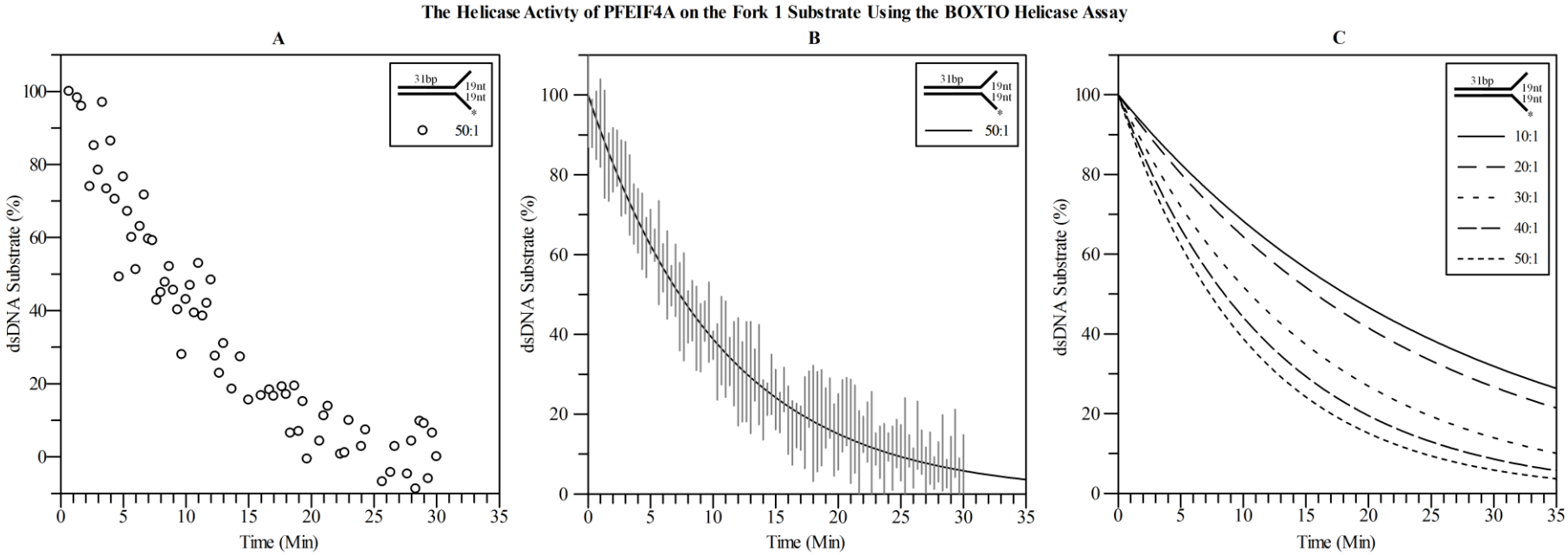


Figure 4.21 **BOXTO Helicase Assay of PFEIF4A Titrations against the Fork 1 Substrate.** Graph A shows a single experimental output obtained from the BOXTO helicase assay at 50:1, PFEIF4A to Fork 1 substrate molar ratio, while graph B shows the line of best fit for the average of three sets of output data with the error bars shown in grey. The data is corrected for any photobleaching, by subtracting its corresponding negative control, and its end point determined from its corresponding positive (boiled) control. Data is fit to a single exponential decay curve. Graph C shows the lines of best fit for the averaged results from a series of titrations of PFEIF4A to the Fork 1 substrate, ranging from 10:1 to 50:1 molar ratios; the solid, wide-spaced dashed, wide-spaced dotted, close-spaced dashed, and the close-spaced dotted line represents their line of best fit. The error bars and a 0:1 ratio experiment, that showed no unwinding, has been omitted for clarity.

and RNA (Pradhan *et al.* 2005), and it was stimulated upon phosphorylation at serine and threonine residues (Pradhan *et al.* 2005). Its helicase activity can be supported by dATP instead of ATP and was inhibited by ATPγS (Pradhan *et al.* 2005; Pradhan *et al.* 2007). Moreover, ATPase and ATP binding was stimulated in the presence of ssDNA (Pradhan *et al.* 2007).

4.2.1 *In silico* Analyses

Like PFEIF4A, PFDH60 was analysed using a variety of *in silco* techniques. An initial search using its codon-optimised DNA sequences (Appendix 9) with the BLAST server (Altschul *et al.* 1990; Johnson *et al.* 2008) revealed sequence identity to multiple DEAD box helicases. Its most similar was ATP-Dependant RNA Helicase from *Theileria parva*, echoing PFEIF4A highest BLAST hit from *T. annulata*. An alignment of PFDH60 sequence with that of ATP-Dependant RNA Helicase (NCBI accession number: XP_762954.1), using ALIGN (Myers *et al.* 1988) (Appendix 10) reveals a sequence identity of 30.8 %, with the conservation mostly located around the core SF2 family motifs (Figure 4.22).

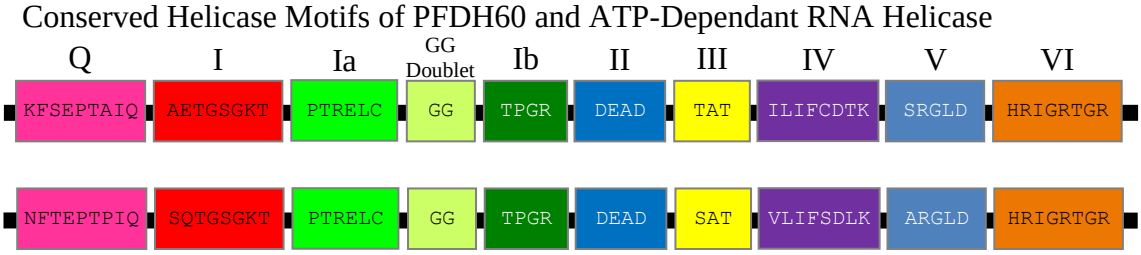


Figure 4.22 **Schematic Based on ALIGN results of PFDH60 and ATP-Dependant RNA Helicase.** Each diagram represents the SF2 motifs present in PFDH60 (top) and ATP-Dependant RNA Helicase (bottom), showing the N-termini on the left and C-termini on the right. Open boxes represent the conserved helicase motifs, and letters inside the boxes are the amino acid sequences of each motif. With only motifs Ia, the GG doublet, II, and VI sharing 100 % identity, there is a lot of sequence variation between PFDH60 and even it most similar BLAST result. Labels above the open boxes are the names assigned to the motifs. The spaces between motifs are arbitrary and not representative of either protein.

Its highest human protein BLAST result was DEAH-box peptide 16 (DHX16; NCBI accession number: CAP58465.1) (Appendix 11), which when compared using ALIGN (Myers *et al.* 1988) reveals a sequence identity of only 13.3 %.

4.2.2 Substrate Specificity Screen

An initial screen for helicase activity was carried out on all 12 substrates shown in Figure 2.6. In short, 20 µL reactions were performed with 1 µM PFDH60 and 100 nM DNA substrate (10:1 ratio). Reactions were incubated at 37 °C for 1 hour and stopped with 5 µL of 5 × termination buffer prior to running on SDS PAGE gels (Figure 4.23). These experiments showed PFDH60 to unwind all of the substrates with no apparent discrimination to substrate topology.

4.3 Discussion

The biochemical analysis described in this chapter has confirmed much of the original characterisation performed by the Tuteja group (Pradhan *et al.* 2005; Pradhan *et al.* 2006; Pradhan *et al.* 2007; Pradhan *et al.* 2007; Pradhan *et al.* 2008). However, it has also expanded and elucidated further information regarding the helicase and ATPase activities of PFEIF4A and PFDH60 which can be used when considering both proteins as novel anti-malarial drug-targets.

4.3.1 PFEIF4A

In silico techniques revealed that PFEIF4A is significantly similar to human EIF4A1 with the core helicase domains and motifs. The biological importance of EIF4A to the eukaryotic cell is well established (Rogers *et al.* 1999; Rogers *et al.* 2001; Rogers *et al.* 2001; Rogers *et al.* 2002; Rozovsky *et al.* 2008) and if PFEIF4A is revealed to be equally vital to *P. falciparum*, it would make an attractive anti-malarial target.

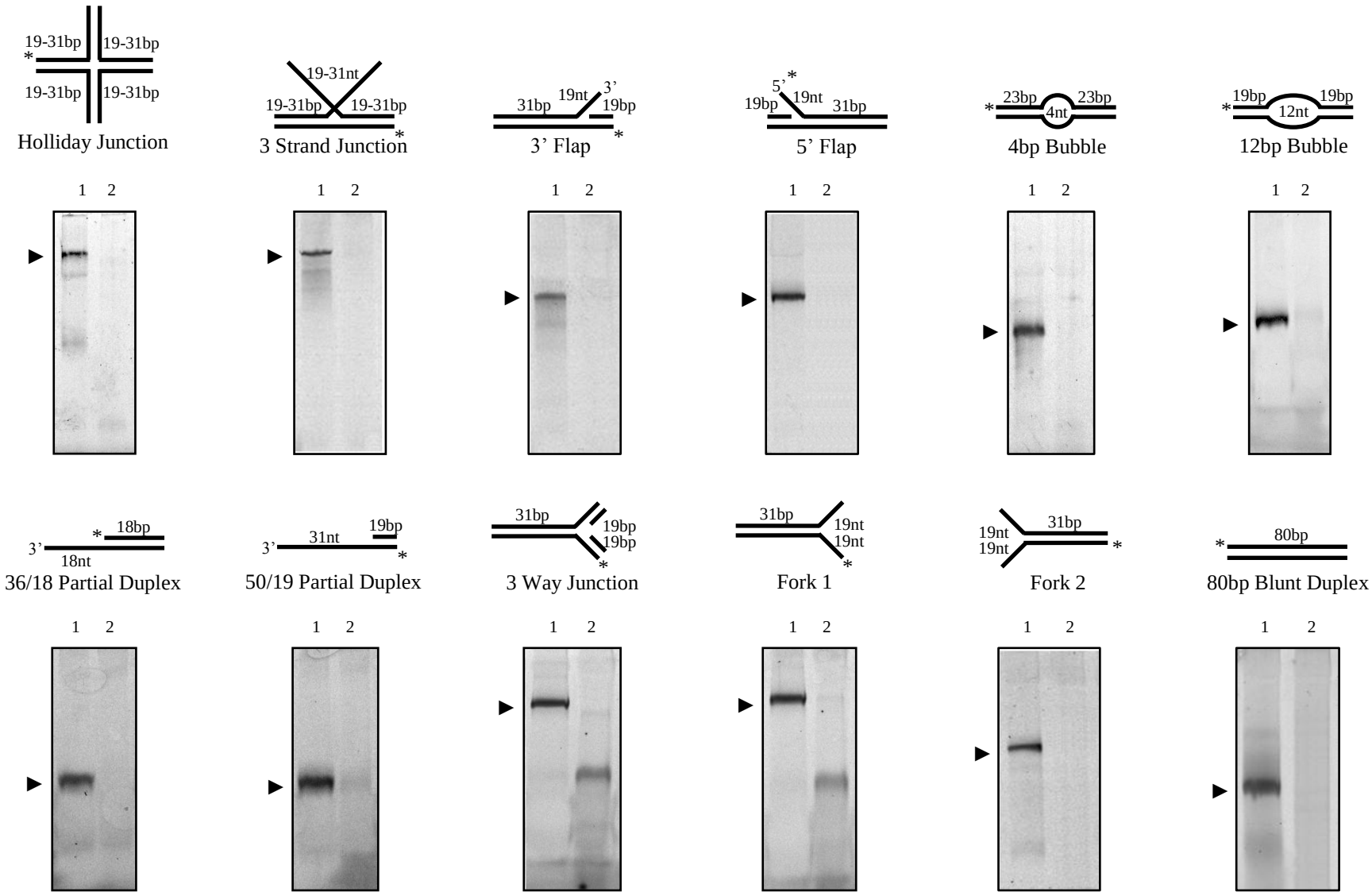

19-31bp
19-31bp
19-31bp
19-31bp
*
Holliday Junction
19-31nt
19-31bp
19-31bp
*
3 Strand Junction
31bp
19nt
3'
19bp
*
3' Flap
5'
*
19bp
19nt
31bp
5' Flap
*
23bp
23bp
4nt
4bp Bubble
19bp
19bp
*
12nt
12bp Bubble
1 2
3'
*
18bp
18nt
36/18 Partial Duplex
3'
31nt
19bp
*
50/19 Partial Duplex
31bp
19bp
19bp
*
3 Way Junction
31bp
19nt
19nt
*
Fork 1
19nt
19nt
31bp
*
Fork 2
*
80bp
80bp Blunt Duplex

Figure 4.23 **PFDH60 Substrate Specificity Screen.** 15 % SDS PAGE gels visualised using a Fujifilm FLA-5000 phosphoimager with a blue laser (473 nm) at 550 V and a green filter (532 nm), showing PFDH60 activity on fluorescein labelled substrates with various topologies as described by Özsoy *et al.* (2003) (Figure 4.3). Above each gel is a representative diagram of the substrate relating to that gel and its associated name as designated by Özsoy *et al.* (2003). The solid black triangle identifies the corresponding band for each of the primary substrates being screened against. Each gel lane is numbered 1-4; lane 1 contains the 37°C control while lane 2 contains the substrate after treating with PFDH60. All of the gels show complete unwinding of the primary substrate after 1 hour at 37°C.

The standard helicase unwinding assay performed by the Tuteja group incorporated a PFEIF4A to substrate ratio of 750:1 with only ~65-70 % of the substrate being unwound after 1 hour at 37 °C (Pradhan *et al.* 2007). However, the work in this chapter with protein from codon-optimised clones has shown complete unwinding of preferential substrates at PFEIF4A to substrate ratios at as low as 30:1 under the same conditions. A possible explanation for the discrepancy was that the Tuteja work was done at the femtomole (fmol) range, while the work presented here was in picomole (pmol) range. Therefore the frequency of interaction would be higher.

The titrations of PFEIF4A to the various substrates elucidated six preferentially unwound substrates. Apparent unwinding at higher concentrations of non-preferential substrates was likely due to opportunistic binding of thermally-frayed ends. Based on the preferential substrates, it appeared that PFEIF4A required a single-stranded overhang, and consistent with Pradhan (2007), this can be either a 3' or 5' overhang. However, the 5' overhang substrate (5' Fork) was not fully unwound after the 1 hour incubation. Interestingly, based on the observations on the 4 bp and 12 bp bubble substrates, a single-stranded region was not adequate to stimulate nucleic acid unwinding, but instead PFEIF4A requires a free single-stranded overhang. The discrepancy between unwinding of Fork 1 and Fork 2 might be explained by the bubble-like structure conferred by small regions of complementary sequence in the Fork 2 free single-stranded ends (Table 4.2). Any observable unwinding may be due to a dynamic equilibrium of forked and bubble-like substrates, with only the former supporting unwinding.

The PFEIF4A Specific Substrate Screen showed that PFEIF4A can resolve DNA substrates containing hairpin structures providing they have a free single-stranded overhang. Similar to the Substrate Specificity Screen and helicase titrations however, the degree of unwinding of substrates with a 3' single-stranded overhang and those of a 5' single-stranded overhang were not equal. The Substrate Polarity Screen, contrary to previous observations made by Pradhan (2007), suggests that only unwinding activities in the 3'-5' direction were stimulated approximately two-fold in the presence of ATP.

The Single-Stranded Overhang Requirements show that PFEIF4A duplex unwinding is supported by the presence of a single nucleotide overhang, albeit slower than for a substrate with a 6 nucleotide overhang. This was likely because one nucleotide was a

smaller ‘target’, affectively decreasing the likelihood of a chance encounter between PFEIF4A and the substrate. The associated increase in unwinding rate with the shortening of the overhang also suggest that PFEIF4A ‘threads’ over the single-stranded overhand, offering credibility to its absolute requirement for a free single-stranded overhang and that a single-stranded region, as exampled by the 4 bp and 12 bp bubble substrates, was not sufficient to support unwinding.

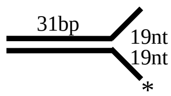
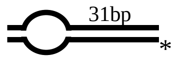
<table>
<tr><th>Constituent Oligonucleotide Sequences</th><th>Revised Structure</th></tr>
<tr><td>5′ GACGCTGCCGAATTCTGGCTTGCTAGGACATCTTTGCCCACGTTGACCCG
   ||||||||||||||||||||||||||||||| : :   :   : :
3′ CTGCGACGGCTTAAGACCGAACGATCCTGTACGACAGATCTCTGATAGCG</td><td> 
</td></tr>
<tr><td>5′ GACGTCATAGACGATTACATTGCTAGGACATGCTGTCTAGAGACTATCGC
   ::::   ::   : |||||||||||||||||||||||||||||||
3′ CTGCGACGGCTTAAGACCGAACGATCCTGTACGACAGATCTCTGATAGCG</td><td> 
</td></tr>
</table>

Table 4.2 **Revised Structure of Fork 2.** The oligonucleotide sequences used to construct the Fork 1 and 2 substrates described by Özsoy (2003) (substrate 10 and 11 in Table 2.3 respectively) were scrutinised using OligoAnalyzer 3.1 (Intergrated DNA Technologies). The predicted heterodimers with the highest Gibbs functions (ΔG), -63.52 and -54.78 kcal mol^{-1} for the Fork 1 and Fork 2 substrates respectively, are shown in the first column. The solid line between the upper and lower sequence represent strong base pair interactions, whereas the dotted lines represent weak base pair interactions. The second column contains the relative schematic for each substrates structure. The structure for Fork 1 is the same as described in Table 2.3; however the structure for Fork 2 is revised based on the heterodimer sequence.

Other work from the Tuteja group has shown PFEIF4A to bind (Pradhan *et al.* 2008) and unwind (Pradhan *et al.* 2007) both DNA and RNA. Likewise, the results presented here have shown that its unwinding activity was unaffected by the incorporation of an RNA strand to form a both DNA:RNA and RNA:DNA hybrid substrates, but curiously the dsRNA substrate was not unwound. Although PFEIF4As activity on other dsRNA substrates or differing sequence and topologies would be beneficial, this observation does present a stark difference between the *P. falciparum* and human forms of EIF4A.

Both the results presented here and in those previously reported by Pradhan (2007) identify PFEIF4As reliance and optimal unwinding activity in the presence of Mg^{2+}. Approximately 20 % of PFEIF4As activity can be recovered by Mn^{2+}, the inability of the other cations studied here to stimulate helicase activity may be due to their increased atomic radii when compared to Mg^{2+}.

In addition, this work has recognised PFEIF4A's broad active-temperature range either side of its 37 °C optimum. However, a decrease in helicase activity was observed at 40 °C. Furthermore, this loss in activity appears to be unrecoverable.

As per Pradhan *et al.* (2008) this work also reports an optimal unwinding activity in the presence of ATP. However, the Tuteja group described dATP, dCTP, and UTP as alternative hydrolysis substitutes, whereas work presented here only identifies dCTP and dGTP as being capable of supporting significant helicase activity. It should be noted that the Tuteja group did not identify any passive (without ATP) activity of PFEIF4A, making comparative observations as to the degree of significant unwinding difficult.

The EMSA experiments show that PFEIF4A has comparative dissociation constants (K_d) for ssDNA (JS80), the 50/19 Partial Duplex, and the Fork 1 substrate. This correlates with the previous observation that PFEIF4A helicase activity has an absolute requirement for a single-stranded overhang. The reported K_d for the dsDNA (80 bp Blunt Duplex) obtained from the GraFit5 software package appears relatively high when compared to a visual reading from the graph (A), which places the K_d between 20 and 25 µM. This was likely due to the incompleteness of the data, where the percentage bound of dsDNA never reaches 100 %, making it harder to forecast an accurate K_d. The reported K_d for the ssRNA was similar to that of the ssDNA, supporting the observation that DNA:RNA and RNA:DNA hybrid substrates were unwound to the same degree as dsDNA substrates of equal topology. However, the reported K_d for the equivalent dsRNA substrate was equally similar despite a lack of observable unwinding.

The hydrolysis of ATP by PFEEIF4A has been reported by Pradhan (2007). However, this work has expanded on this to show nucleic acid presence stimulates PFEIF4As ATPase activity, and that despite preferential substrates for unwinding, DNA substrate topology does not affect it. The apparent background ATP turnover suggested by the no DNA control may be due to residues with hydrophobic side chains, as hydrophobic groups have

been shown to interact with ammonium molybdate causing it to crash out of solution giving false positives (Feng *et al.* 2011). Arginine (R), Asparagine (N), Glutamine (Q), and Lysine (K) make up 5.6, 2.8, 4.6, and 6.7 % of PFEIF4A's amino acid composition and all contain hydrophobic side chains.

When considering that the malachite green assay was performed at ten-fold higher concentrations of PFEIF4A, DNA substrates, and ATP to those used during the luciferase ATPase assays, they gave proportionally similar results regarding DNA topology. The initial rates taken from the luciferase assay were slightly higher and background ATP turnover was not apparent. This was likely due to the luciferase assay reportedly being a more sensitive technique. The NADH ATPase assay did not yield any discernible data.

The triplex displacement assays did not show any apparent displacement of the TFO on either of the substrates. This could be due to PFEIF4A not being capable of translocating dsDNA, or that due to its high K_d for dsDNA when it encounters the TFO it simply dissociates from the duplex leaving the TFO bound. Further experiments at increasing concentrations of PFEIF4A might show displacement of the TFO, offering weight to the latter theory.

BOXTO concentration used in the high-throughput assay was similar to the 3:1 BOXTO to dsDNA optimum ratio described by Ahmad (2007). The ability to discriminate between ssDNA and dsDNA at a 3:1 was not largely improved, and a 5:1 ratio gave a greater overall fluorescence output. It was for these reasons that a 5:1 was used to perform the assay thereafter.

The assay itself showed comparable amounts of unwinding to those obtained from the gel based assay, with both the assays showing 100 % of the substrate being unwound at approximately 30 minutes at a molar ratio of 50:1, PFEIF4A to the Fork 1 substrate. However, the initial rates obtained over the first 15 minutes from each method are notably different; 74.49 ± 2.76 fmol min^{-1} for the gel based assay, and 113.18 ± 3.93 fmol min^{-1} for the BOXTO based assay. The difference is likely due to the better resolution of the BOXTO based unwinding.

Based on the Singleton *et al.* (2007) mode of classification, PFEIF4A is a SF2Aα helicase due to its 3'-5' polarity bias (A) and its absolute requirement for a single-stranded overhang, indicating a passive helicase mechanism (α).

The unwinding rate of PFEIF4A is ~1 bp s^{-1} based on the 50/19 Partial Duplex substrate (substrate 23, Table 2.3). However, this is assuming that the rate of translocation of PFEIF4A on ssDNA is the same as its dsDNA unwinding rate. If this were the case then the rate, in terms of base pairs per second (bp s^{-1}), would have been the same for all of the Partial Duplex substrates of different single-stranded overhang length. Instead the rate of unwinding the 25/19 Partial Duplex substrate (substrate 29, Table 2.3) is ~2.5 bp s^{-1}. This could be due to a mechanical difference in the translocation of PFEIF4A along ssDNA compared to the unwinding of dsDNA. Alternatively, it may be that as the single-stranded overhang length decreased, so its rigidity increased (Chen *et al.* 2012), presenting a more consistent topology for a chance encounter with PFEIF4A.

4.3.2 PFDH60

In silico techniques reveal a stark difference in sequence with related human proteins, potentially leading to a greater chance of discrimination when considering exploiting PFDH60 as an anti-malarial drug target.

Unlike PFEIF4A, PFDH60 showed no discrimination between substrates during the Substrate Specificity Screen. Furthermore, all of the substrates were unwound at a 10:1 PFDH60 to substrate molar ratio (1 µM to 100 nM).

Unfortunately further experiments could not be performed, initially due to PFDH60s storage issues then due to time restraints. However, based on the experimentation that was performed, PFDH60 can be identified as a SF2β helicase, due its ability to unwind blunt-ended dsDNA using an active helicase mechanism (Singleton *et al.* 2007).

CHAPTER FIVE: STRUCTURAL CHARACTERISATION

5.1 Introduction

Several structural-based studies (Korolev *et al.* 1997; Velankar *et al.* 1999) have elucidated helicase structure-function mechanisms. When combined with biochemical data, this has lead to proposed models for duplex unwinding (Soultanas *et al.* 2000; Soultanas *et al.* 2001). Furthermore, the identification of interactions between specific residues and the substrate has been a driving force behind mutagenic studies on the conserved helicase motifs (Pause *et al.* 1992; Korolev *et al.* 1998; Mahdi *et al.* 2003; Tanner *et al.* 2003), many of which have lead to observations of both increased (Dutta *et al.* 2011) and decreased (García-Gómez *et al.* 2011) activities.

The relatively recent field of structural-guided drug design also highlights the potential application of solved protein target structures which can be used to elucidate sequential and topological features for drug design and inhibition. Several crystal structures exist with the incorporation of bound inhibitors to clinically important proteins. These include Human Immunodeficiency Virus (HIV) 1 protease (Blum *et al.* 2011), dihydropteroate synthase from the gram-negative bacterium *Burkholderia cenocepacia* (Morgan *et al.* 2011), and trypanothione reductase (TRYR) from the *Trypanosoma brucei* parasite (Patterson *et al.* 2011). High resolution Nuclear Magnet Resonance (NMR) data has also been used to perform *in silico* ligand and structural-based drug design (Hajduk *et al.* 2005). A study performed by Chen *et al.* (2009) expanded on the original Hepatitis C virus (HCV) helicase domain crystal structure (Yao *et al.* 1997), by crystallising an inhibitor bound protein which lead to the design of more potent inhibitors that suppress HCV replication.

Studies highlighting the usefulness of known structures for *in silico* screening have been published (Kaczor *et al.* 2010), which in turn has lead to *in silico* based methodologies being devised such as Active Site Pressurisation (ASP) (Withers *et al.* 2008).

This chapter will describe preliminary structural work on PFEIF4A and PFDH60 which can be used to provide a firm basis for future structural experimentation.

5.2 PFEIF4A

At present only a structural comparison between the theoretical N- and C-core domains of PFEIF4A with human EIF4A3 has been published by the Tuteja group (Pradhan *et al.* 2007).

5.2.1 *In silico* Analysis

The FASTA protein sequence of PFEIF4A (Appendix 7) was analysed using the Protein Homology/Analogy Recognition Engine (PHYRE) server version 2 (Kelley *et al.* 2009). This compares the sequence of interest with up to one-thousand protein sequences with known structures, with a priority for proteins with the lowest e-values based on their sequence alignment to PFEIF4A to generate a theoretical structure. The resulting Protein Database (PDB) file can then be visualised and annotated using a molecular imaging program (Figure 5.1).

The resulting structure predicted for PFEIF4A showed two domains. These resembled the N- and C-core domains and showed the core helicase motifs lining the cleft between the two. It also showed a high level of superimposition with the human EIF4A isoform 1 structure.

5.2.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

All NMR experiments were kindly performed by Dr Chris Read (University of Portsmouth, Hampshire, UK). NMR spectroscopy was employed to evaluate the proteins' suitability for more in depth structural studies. 1D NMR spectra showed multiple peaks suggesting some degree of tertiary structure; this was followed by multiple 2D Heteronuclear Single Quantum Coherence (HSQC) experiments (Bodenhausen *et al.* 1980) (Figure 5.2) on ^{15}N-labelled full-length His_6-T-PFEIF4A and its two subdomains. The subdomains were prepared by size exclusion chromatography of the degradation products from full-length His_6-T-PFEIF4A as described under the storage conditions in chapter three.

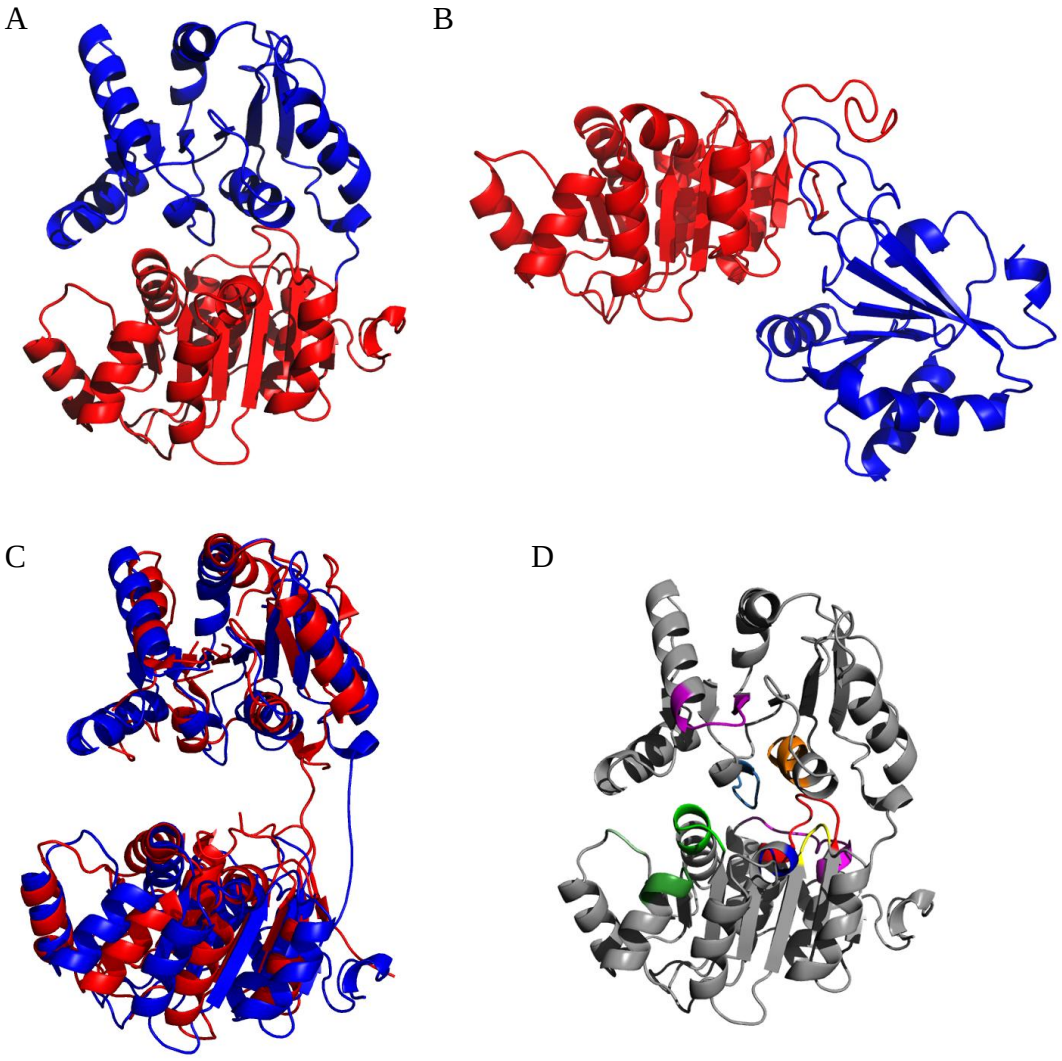


Figure 5.1 **Structural Prediction of PFEIF4A.** Rendered using PyMol v0.99; render A shows the standard PHYRE2 server predicted structure of PFEIF4A orientated with the N-core domain at the bottom (red) and the C-core domain at the top (blue). Render B shows the intensive PHYRE2 server predicted structure with its N-core domain (red) orientated to match that of the standard structure prediction. Its C-core domain (blue) is not only orientated around its N-core differently, it is also structured dramatically differently. Render C shows the standard structure predicted in blue overlaid with Human EIF4A1 (PDB code; 2ZU6) shown in red. Some manipulation of the non-structured linker region was required to overlay both domains. Due to the high level of superimposition between the homologous domains, the standard structure prediction was used to annotate the core helicase motifs (D). The motif colour scheme of render D follows that of Figure 4.1; Q motif is in pink, motif I, Ia, and Ib are in red, bright green, and dark green respectively, the GG doublet in pale green, motif II in bright blue, motif III in yellow, motif IV in purple, motif V in pale blue, and motif VI in orange.

The initial spectra obtained from the 1D NMR showed multiple well defined peaks corresponding to chemical shifts associated with a folded protein. However, 2D NMR on the full-length protein and the N- and C-terminal domains also gave poorly resolved spectra. Due to the time associated collecting the NMR data and the priority focus of this study, no further NMR experiments were carried out.

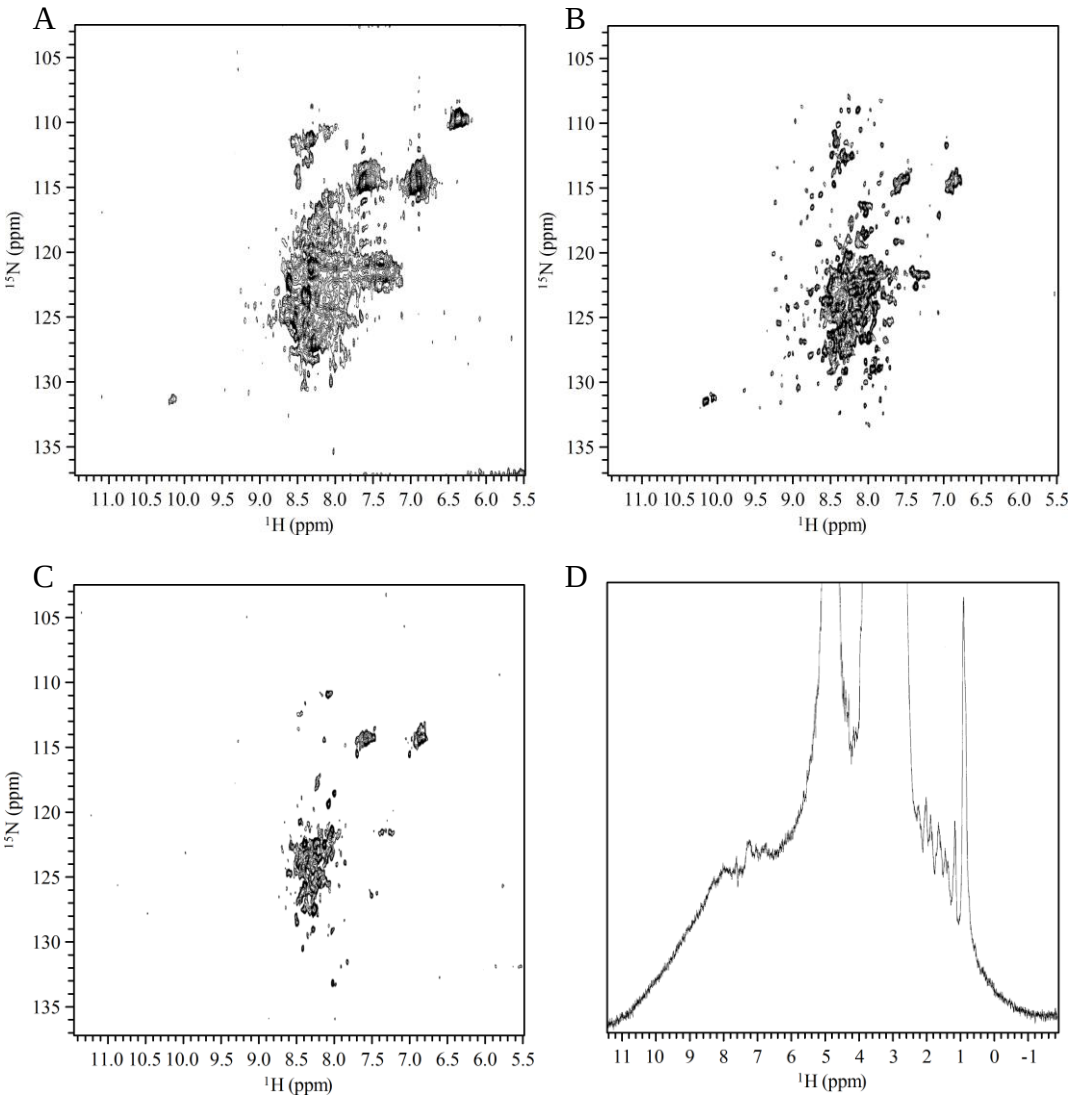


Figure 5.2 **PFEIF4A NMR Experiments.** Panel A shows the 2D spectrum from a HSQC experiment on a ^{15}N-labelled full-length His_6-T-PFEIF4A, while panels B and C show the spectra of its N- and C-terminal domains. After personal communication with Dr Chris Read, the pH was reduced in an attempt to increase resolution. However, resolution was decreased and the poorly defined peaks at frequencies 5.5 to 9 parts-per-million (ppm), on the 1D spectrum of the N-terminus (panel D) suggest a partially unfolded state.

5.2.3 Analytical Ultracentrifugation (AUC)

Sedimentation AUC was performed on a 0.6 mg mL^{-1} sample of His_6-T-PFEIF4A, correlating to 1 Absorbance Unit (AU) at 280 nm, based on PFEIF4As molar extinction coefficient of 29715 M^{-1} cm^{-1} calculated with ProtParam (Wilkins *et al.* 1999), for 100 scans with 3 minute intervals at 45500 × g at 20 °C, based on the anticipated molecular weight (Chervenka 1970). The data collected was analysed using SEDFIT (Schuck 2000) to give some insight into His_6-T-PFEIF4As size and shape (Figure 5.3).

Based on the sedimentation coefficient (s = 3.25) and the molecular weight (M_r = 48853) obtained from the AUC run, a frictional coefficient ratio (f/f_0) of 1.45 and a diffusion coefficient of 6.116 × 10^{-07} cm^2 s^{-1} were calculated. His_6-T-PFEIF4As hydrodynamic radius (R_h) and rotational radius (R_0) were calculated as 3.50 and of 2.42 nm respectively. Based on the frictional coefficient ratio, axial ratios of 6.12 for an oblate ellipsoid model and 5.63 for a prolate ellipsoid model were also calculated.

5.2.4 Small Angle X-ray Scattering (SAXS)

Raw SAXS data was kindly collected by Dr Anastasia Callaghan (University of Portsmouth, Hampshire, UK). Scattering patterns were taken from successive 90 second exposure times. These were used to check for radiation-induced damage or aggregation of the protein and were performed on a range of His_6-T-PFEIF4A concentrations from 1 to 10 mg mL^{-1}. The data was normalized for experimental noise, including that from the intensity of the incident beam and any scatter observed when analysing the blank solvent.

Using the PRIMUS program (Konarev *et al.* 2003) from the ATSAS software suite (Konarev *et al.* 2006), the sample data was first normalised against the blank buffer data, then translated into a Guinier approximation to elucidate information on His_6-T-PFEIF4As' radius of gyration (R_g), which was obtained from the slope of the Guinier plot. The same normalised data was then transformed to a Kratky plot, where a bell-shaped curve indicates a globular macromolecules folded state based on Porod's law (Putnam *et al.* 2007). Using the GNOM program (Svergun 1992), also from the ATSAS software,

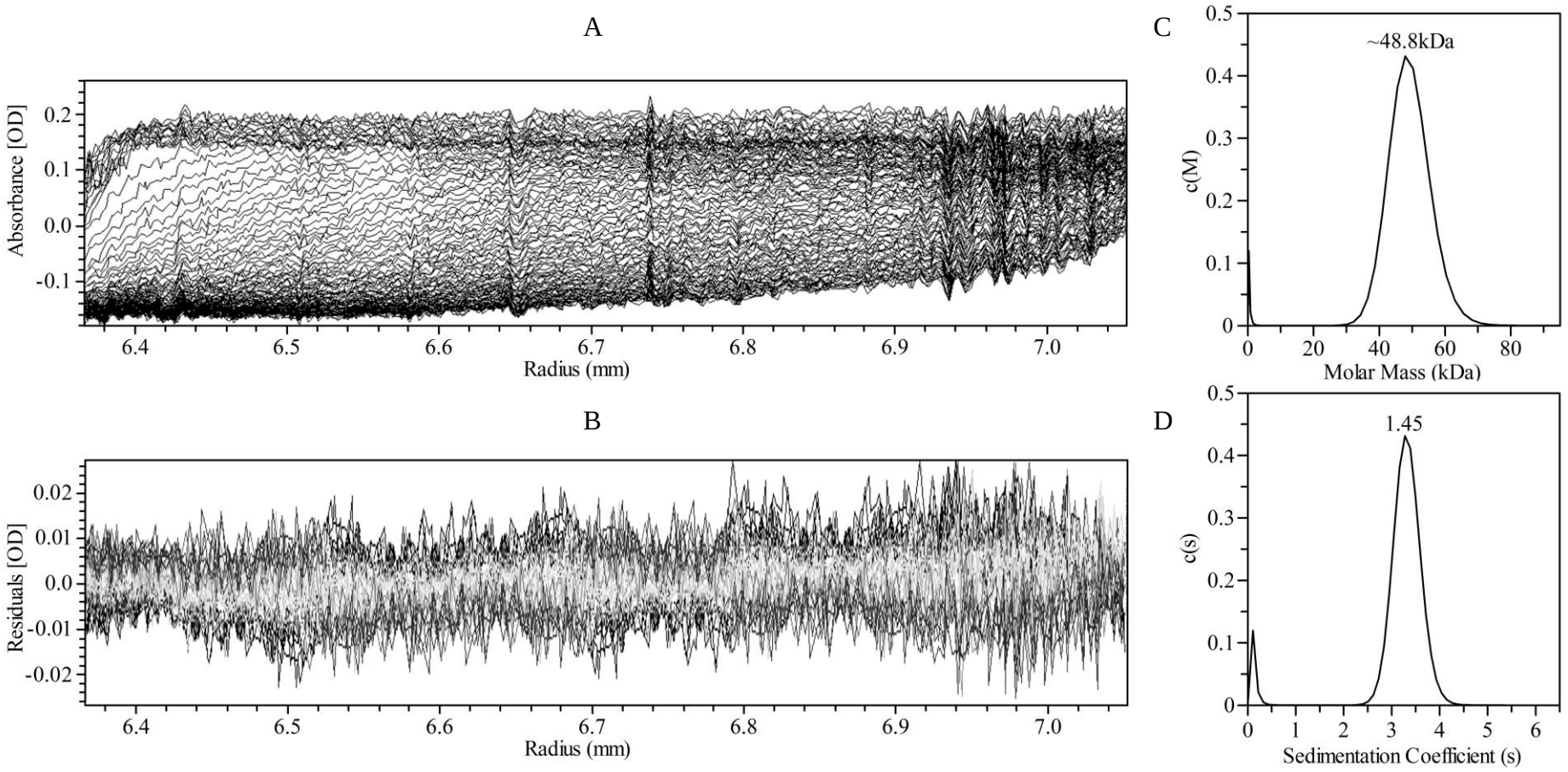


Figure 5.3 **AUC Data Analysis of PFEIF4A at 45500 × g.** Panel A shows the residuals plot taken from between the proximal and distal menisci of the raw data with the radius of the ultracentrifugation cell along the *x*-axis and absorbance along the *y*-axis, together with residuals for the fit in panel B. Panel C shows the continuous c(M) distribution obtained from the analysis of the data from A and B, estimating the molar mass of PFEIF4A. Panel D shows the continuous c(s) distribution also calculated from panels A and B, estimating PFEIF4As sedimentation coefficient. Both traces are annotated with their peak values.

pair-distribution functions, P(r), were generated (Figure 5.4) which could subsequently be used to estimate the particle maximum dimension (D_{max}) and calculate the real-space R_g value.

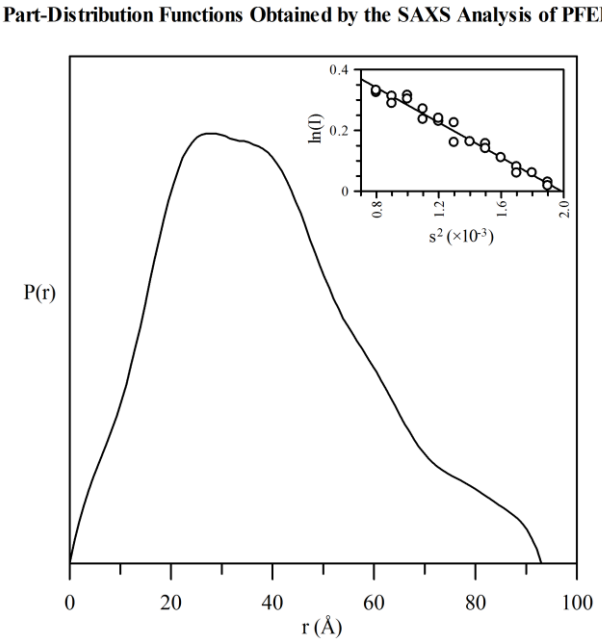


Figure 5.4 **Pair-Distribution Functions of PFEIF4A.** The trace of the P(r) plot was generated using the Guinier region shown in the inset panel. The slightly bilobal shape lends some suggestion to a two domain structure. The radius in angstroms (Å) is shown on the *x*-axis. The *x* intercept indicates PFEIF4As largest dimension (D_{max}) is approximately 88 Å (8.8 nm).

The P(r) data was used to generate multiple low resolution reconstructions using the DAMMIF program (Svergun 1999; Franke *et al.* 2009) under its default parameters. Then, using the DAMAVER program (Svergun 1999), ten *ab initio* reconstructions were aligned, averaged and filtered for areas of low density (Figure 5.5). Increasing the number of aligned reconstructions did not significantly alter the averaged model. Finally, the DAMAVER model was overlaid by the PHYRE2 structure described in Figure 5.1 using the SUPCOMB program (Kozin *et al.* 2001) (Figure 5.6). This showed that the DAMAVER model could comfortably accommodate the PHYRE2 structure.

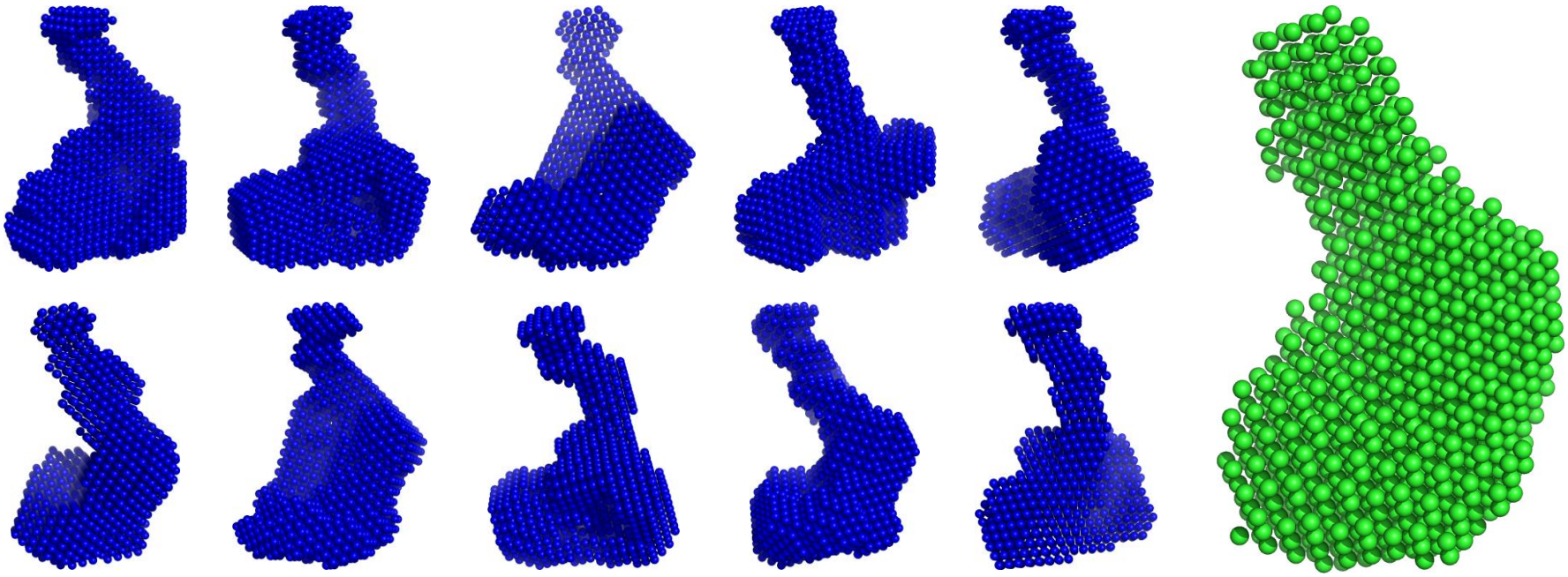


Figure 5.5 ***Ab Initio* Reconstructions and Averaged SAXS Model of PFEIF4A.** The 10 *ab initio* reconstructions obtained from the DAMMIF software (Franke *et al.* 2009) are shown in blue. Each model has been presented to illustrate their resemblance to each other, albeit arbitrary. These models were averaged and filtered for areas of low density using the DAMAVER software (Svergun 1999) and the resulting model is shown in green. Again, its orientation serves only for its comparison to the *ab initio* models.

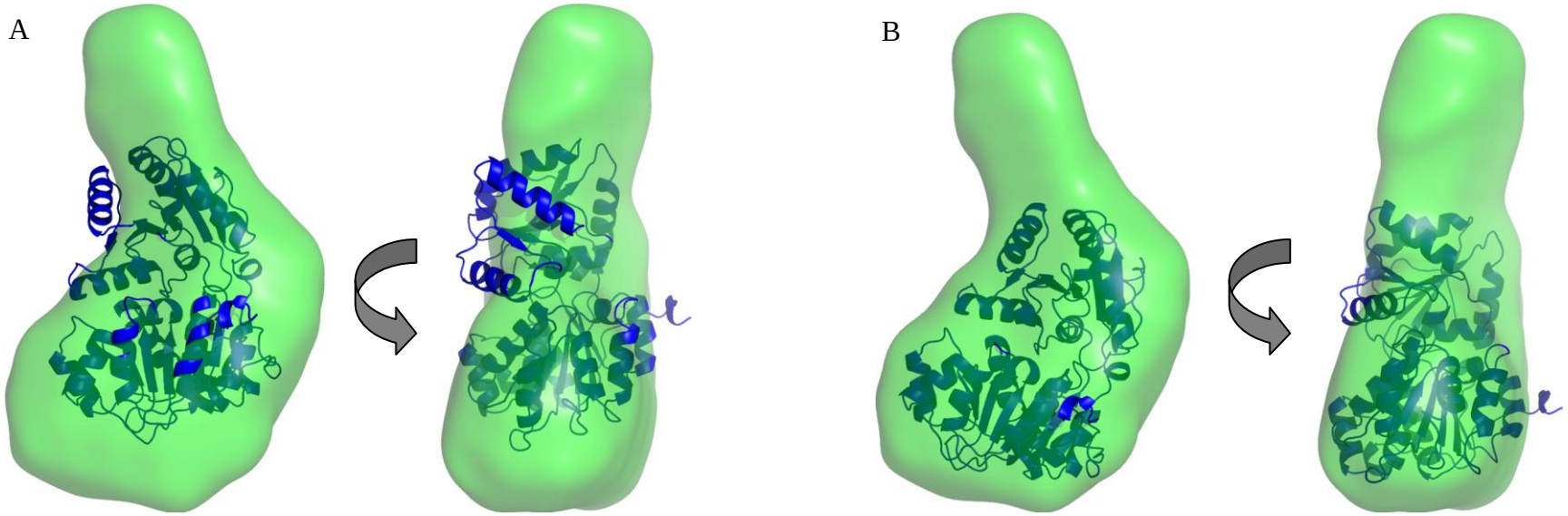


Figure 5.6 **Superimposition of SAXS Generated Model with PHYRE2 Server Model of PFEIF4A.** Panel A shows the SUPCOMB (Kozin *et al.* 2001) derived superimposition of the PHYRE2 server-generated PFEIF4A model (blue) over a surface render of the SAXS model (green) generated by DAMAVER (Svergun 1999). The fit is clearly poor, with the PHYRE2 server model appearing to have been overlaid at the *y*-axis median, giving a normalised spatial discrepancy (NSD) of 3.16. However, manual overlay (B) of the PHYRE2 server model does illustrate the potential for a more convincing fit. The large protuberance shown at the top of the DAMAVER model, possibly conferred by the addition amino acids sequence associated with the intact His_6-tag of the SAXS sample; might be responsible for skewing the superimposition of the PHYRE2 server model, hence the poor fit. In both panels A and B, the right-most images are the orthogonal view of the images to their left following a 90° rotation around their *y*-axis.

5.2.5 Dynamic Light Scattering (DLS)

Based on the observation in chapter four that helicase activity was dramatically reduced at incubation temperatures of 40 to 45 °C, DLS experiments were employed to observe any temperature-induced structural changes of His_6-T-PFEIF4A. A 100 µL sample was prepared at 0.6 mg mL^{-1}, correlating to 1 AU based on PFEIF4As molar extinction coefficient. The values for R_h and scatter intensity were monitored during 10 minute incubations at various temperatures between 20 and 65 °C. Due to the temperature effects on viscosity, each measurement was corrected for background fluctuations by scanning a blank buffer sample at each of the respective temperatures.

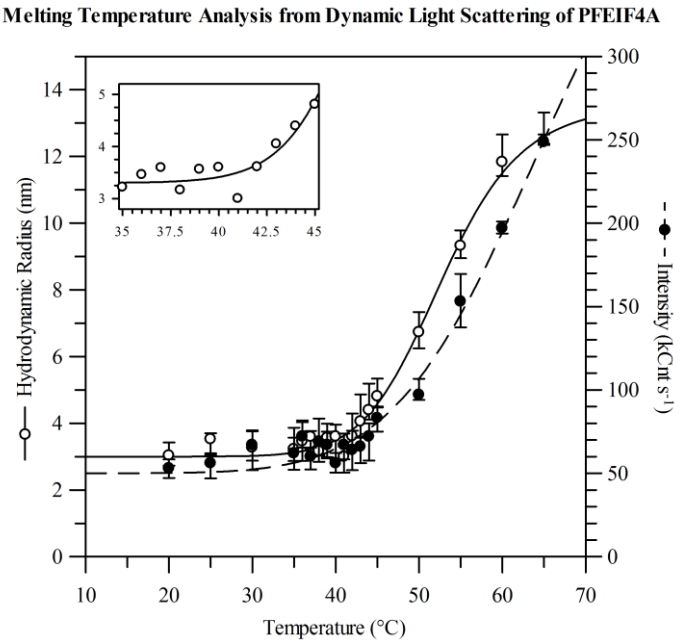


Figure 5.7 **Effect of Temperature on PFEIF4As Structure.** The graph shows the change in hydrodynamic radius (R_h), shown as white circles with a solid line of best fit, and the intensity of scattering, shown as black circles with a wide-spaced dashed line of best fit, of PFEIF4A with increasing temperature. The inset graph shows the magnified portion of the R_h between 35 and 45 °C, highlighting the onset of increasing R_h at 43 °C.

As the temperature increased, both R_h and scattering intensity increased sharply at 43 °C (Figure 5.7). Furthermore, a sample incubated at 65 °C then subsequently cooled to 30 °C did not give the same R_h or scatter intensity as at 30 °C, suggesting an irreversible secondary or tertiary structural change.

5.3 PFDH60

Other than the identification of the core helicase motifs of PFDH60 from basic bioinformatic techniques (Pradhan *et al.* 2005), no structural characterisation has been performed on PFDH60.

5.3.1 *In silico* Analysis

The FASTA protein sequence of PFDH60 (Appendix 9) was analysed using the PHYRE2 server (Kelley *et al.* 2009) generating a theoretical structure which was then annotated (Figure 5.8).

The standard structure prediction omitted the first 278 and the last 6 amino acid residues, as well as those between 363 and 431, whereas the intensive structure prediction included every residue. However those omitted from the standard prediction contributed to large amounts of unstructured regions within the intensive structure. In spite of this, both structures predicted for PFDH60 showed two domains. These resembled the N- and C-core domains and the core helicase motifs were shown lining the cleft between the two.

5.3 Discussion

The work described in this chapter provides preliminary insight into PFEIF4A and PFDH60 structures. Coupling the structures with information elucidated from the rest of this study supports a more in depth understanding of their activities, and ultimately their suitability as novel anti-malarial drug targets.

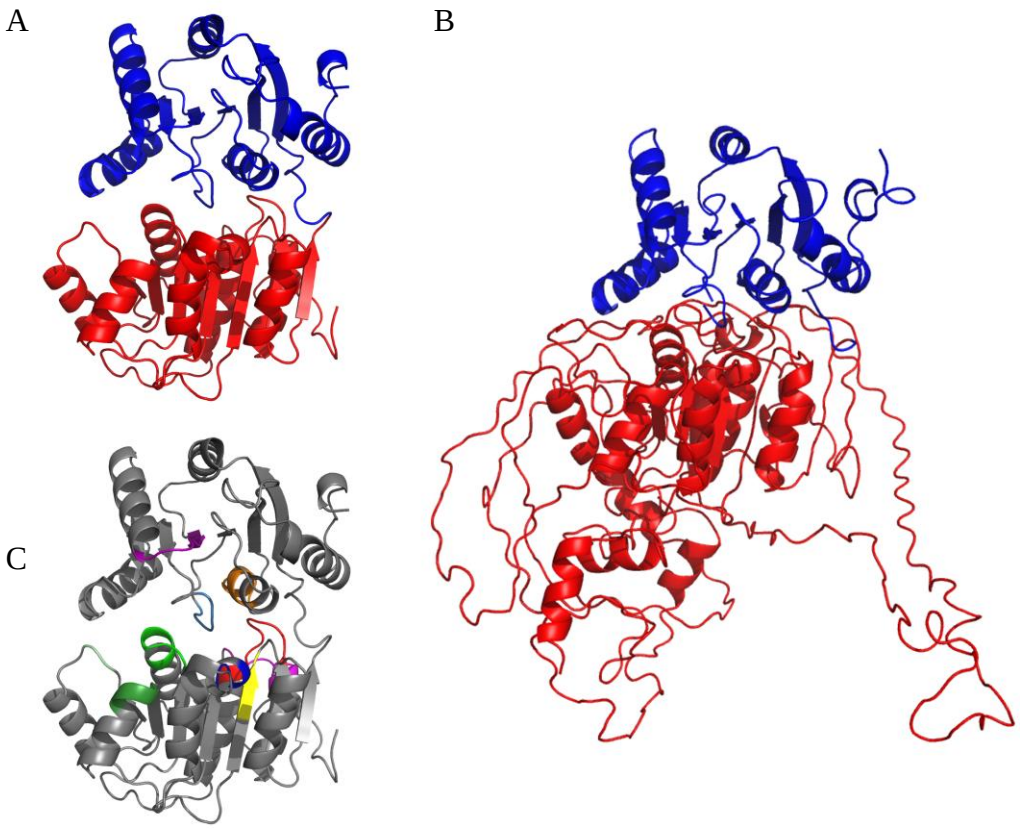


Figure 5.8 **Structural Prediction of PFDH60.** Rendered using PyMol v0.99; render A shows the standard PHYRE2 server predicted structure of PFDH60 orientated with the N-core domain at the bottom (red) and the C-core domain at the top (blue). Render B shows the intensive PHYRE2 server predicted structure orientated to match that of the standard structure prediction. Unlike the standard predicted structure it has a large amount of unstructured regions in its N-core domain (red). For clarity, the standard structure prediction was used to annotate the core helicase motifs (C). The motif colour scheme of render C follows that of Figure 4.1; Q motif is in pink, motif I, Ia, and Ib are in red, bright green, and dark green respectively, the GG doublet in pale green, motif II in bright blue, motif III in yellow, motif IV in purple, motif V in pale blue, and motif VI in orange.

5.3.1 PFEIF4A

The structural comparison performed by the Tuteja group was on the human EIF4A3 (Pradhan *et al.* 2007). However, work presented in this works has already identified that PFEIF4A shares the highest identity with EIF4A1 (Table 4.1), justifying its use for a comparison in Figure 5.1. Furthermore, unlike the comparison made here, the Tuteja group

did not use the full-length protein. A further comparison exists on the International Centre for Genetic Engineering and Biotechnology (ICGEN) bioinformatics website on comparative modelling of *P. falciparum* annotated proteins (Gowthaman *et al.* 2005). However, this model only predicts one domain and it is based on the yeast EIF4A orthologue (1FUU).

The NMR HSQC experiments did not yield high resolution data for either the full-length or the subdomains of His_6-T-PFEIF4A. This is possibly due to His_6-T-PFEIF4A being in the upper-range of the generally accepted size limitation of NMR (Yu 1999). Under the advice of Dr Chris Read, changing the chemical environment by lowering pH did nothing to increase the resolution. Instead it decreased, likely due to the denaturation of the protein. However the data was enough to show different chemical environments suggesting the protein exhibited secondary and tertiary structure.

The relative molecular weight (M_r) obtained from the AUC run is approximately 1.1 kDa smaller than that predicted using ProtParam. Other than any error incurred by either the prediction or AUC run, this discrepancy could also be partly due to a loss of the His_6-tag (6 × 155.15 Da). This would correlate with the weaker western blot detection of intact His_6-T-PFEIF4A with anti-His antibody over-time (as described in 3.2.8). The R_h obtained from the AUC is the theoretical radius of a sphere with the same diffusion rate of His_6-T-PFEIF4A (Schuck 2000; Schuck 2010), and was determined to be 3.50 nm. However, the experimentally determined R_0 was 2.42 nm. The R_0 is the distance from the centre of His_6-T-PFEIF4A to the average extremity of the molecule. Therefore if a molecule's surface has a number of large protuberances, these are cancelled out when averaging against the more concave surfaces, giving a smaller measured radius than predicted from the R_h.

This suggests that the overall shape of the His_6-T-PFEIF4A molecule is irregular, which is substantiated by observations made on the axial ratios. If either axial ratio equals 1 then that dimension is perfectly circular. If it is greater than 1 then that dimension is more ellipsoid. If both the oblate and prolate axial ratios equal 1, the molecule is spherical. However, if one is greater than 1 it would be 'cylindrical'. Both the oblate and prolate axial ratios of His_6-T-PFEIF4A are greater than 1, 6.12 and 5.63 respectively, indicating an irregular shape. Additionally, a f/f_0 over 1.2 also indicates a more globular shape (Schuck 2005). It should be noted however, that the presence of the relatively unstructured His_6-tag may impart a significant hydrodynamic friction, and that the true molecular shape may

exhibit a lower geometric asymmetry. Additionally, using the Hydropro program (Ortega *et al.* 2011) to analyse the PHYRE2 structure, gave an S value of 3.57 compared to the experimentally determined S value of 3.25 from the AUC run. This offers some credibility to the predicted PHYRE2 structure of PFEIF4A.

The normalised spatial discrepancy (NSD) value of 3.16 obtained from the SUPCOMB superimposition of the His_6-T-PFEIF4A PHYRE2 structure over the DAMAVER model indicates an unacceptable fit. The superimposition of identical objects would give an NSD of 0. Superimposed objects that are symmetrically different would give an NSD of greater than 0 and a fit is considered acceptable with NSDs equal to or less than 1. Anything over an NSD of 1.5 would require attention to obtain a better fit. The unacceptable NSD for His_6-T-PFEIF4A is likely due to the presence of the His-tag moiety in the DAMAVER model which is not present in the PHYRE2 structure. Furthermore, taking into account that SAXS is a liquid state structural technique, the data generated is affected by the detection of multiple consistent conformational states of any dynamic regions within the protein, that when averaged can create a model illustrating the protein at its less compact and therefore largest condition. Despite this, other parameters obtained from the analysis of the SAXS data couples with those elucidated by alternative means.

For example, the I(0) values obtained from His_6-T-PFEIF4A and BSA control samples of comparable stoichiometry can be used to estimate the M_r as describe below.

$$\frac{\mathrm{I(0)sample}}{\mathrm{I(0)reference}} \times \text{reference } \mathrm{M_r}$$

The I(0) values obtained from the Guinier plots of a 6.35 mg mL^{-1} sample of His_6-T-PFEIF4A (I(0) = 2.5192 ± 0.0498) and a 6.29 mg mL^{-1} BSA reference (I(0) = 3.6355 ± 0.046, Mr = 66.5kDa), estimates His_6-T-PFEIF4A M_r to be ~46 kDa ± 1.5. This value is similar to the M_r that was obtained by sedimentation AUC (~48 kDa) suggesting the SAXS data itself is accurate.

The data obtained from the DLS experiments revealed a sharp incline in R_h and scattering intensity with increasing temperature at all concentrations. This is due to protein structure being stabilised by hydrogen bonds, hydrophobic interactions, and Van der Waal forces, each of which contributes a small degree of stability to the overall structure. These

stabilising forces can be disrupted by the addition of free energy, such as an increase in temperature, potentially denaturing the protein into a less ordered or unfolded state. This results in a less compact structure which effectively increases the average R_h of the sample. In addition, the hydrophobic residues once buried within the interior of the folded structure are now exposed to the solvent. Due to the entropically unfavourable state, the hydrophobic residues from one denatured protein may associate with those of another. This effectively increases the molecular weight of the sample protein which in turn increases the scattering intensity. The temperature at which half the maximal unfolding of a protein occurs is defined as the melting point (T_m), placing the T_m of PFEIF4A at approximately 52.5 °C.

Although these structural studies have elucidated limited information on both helicases, structural studies have emerged as an effective way to guide drug and inhibitor design. Therefore, this may provide a basis for future work due to the increased incorporation of such studies in clinical approaches.

5.3.2 PFDH60

Due to the issues associated with storage of PFDH60, no experimental analysis could be performed. However, the PHYRE2 server predicted model showed large amounts of unstructured regions. This may explain why PFDH60 loses activity in a relatively short amount of time.

CHAPTER SIX: INHIBITOR CHARACTERISATION

6.1 Introduction

With over one million deaths annually (Greenwood *et al.* 2005), malaria poses a huge threat to some of the world's most vulnerable people and its long-term control and eradication is presently unforeseeable (Tuteja 2007). The present strategy in combating malaria, promoted by the World Health Organisation (WHO), involves a combination of vector control strategies and parasitic interventions (WHO 2008; WHO 2009; WHO 2010). For many countries, treatment consists of either chloroquine or sulphadoxine-pyrimethamine, or a combination of the two, however wide-spread resistance has long been identified (EANMAT 2003; Plowe *et al.* 2004). This led to a search for novel ways to combat the disease and, following the completion of the *P. falciparum* genome (Gardner *et al.* 2002), a surge in bioinformatics inspired numerous novel approaches in combating the disease.

Many of the emerged strategies have involved the targeting of the parasite's proteins, such as those implicated during gametocytogenesis (Baker 2010; Kuehn *et al.* 2010) and epigenetic modification (Salcedo-Amaya *et al.* 2010). As previously discussed, helicases have been suggested as effective anti-malarial drug-targets (Tuteja 2007) due to their involvement in essential cellular processes. The *Plasmodium* proteome is an attractive target as only approximately 40 % of its genome has been annotated, offering considerable potential for the discovery of new anti-malarial drug-targets (Birkholtz *et al.* 2008).

Previous studies have shown that *Plasmodium falciparum* growth was successfully inhibited with PFEIF4A antibodies, and dsRNA in RNA interference experiments (Pradhan *et al.* 2007). Experiments into the effects of DNA binding compounds on PFEIF4A unwinding and ATPase activities have also been performed (Pradhan *et al.* 2008). These showed that out of a variety of well-known DNA binding compounds, actinomycin, 4',6-diamidino-2- phenylindole (DAPI), daunorubicin, ethidium bromide, netropsin and noglamycin had inhibitory affects. These observations supported the suggestion that PFEIF4A would make a viable anti-malarial drug target (Tuteja 2007).

This chapter will describe the effect of a novel DNA binding drug on helicase and translocase proteins, as well as exploring the potential of DNA aptamers in identifying critical binding sites for the development of novel drugs.

6.2 GWL-78

The work discussed in this section was performed alongside Mr Irfaan Dawood as part of collaboration with Prof. David Thurston (King's College London, London, UK) who kindly provided a DNA-interactive C8-*bis*-pyrrole pyrrolobenzodiazepine (PBD) conjugate called GWL-78.

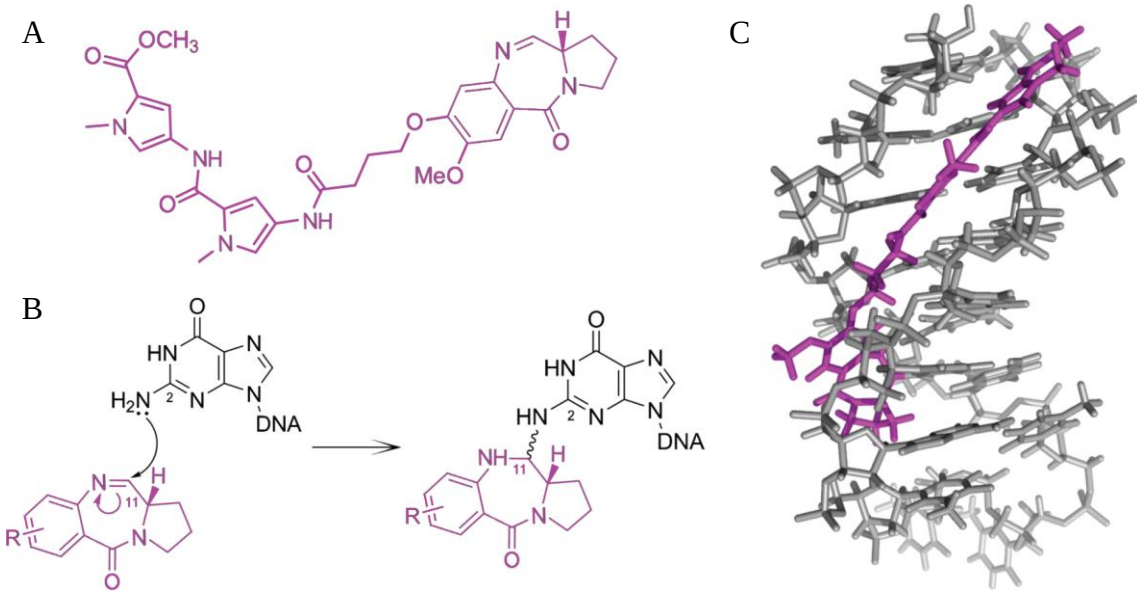


Figure 6.1 **Structure and Binding of GWL-78 to DNA Adapted from Rahman *et al.* (2011).** Schematic A shows the structure of the C8-*bis*-pyrrole PDB conjugate while B details the theorised mechanism of covalent binding of the PDB moiety (magenta) through its C11 to the C2-NH_2 of a guanine in the DNA minor groove. The energy minimised molecular model in C shows the C8-*bis*-pyrrole side chain oriented to the 5'-end of a DNA duplex.

GWL-78 reversibly binds through its C11-position at the C2-NH_2 of guanine bases in the minor groove, with a preference for 5'-XGX(W)z motifs (X is any base, but commonly a purine, W is an A or T, and z is 3 ± 1) (Wells *et al.* 2006; Rahman *et al.* 2011) (Figure 6.1). Furthermore, it has been shown to inhibit the binding of Nuclear Factor Y (NF-Y) to its

target CCAAT box sequence within the human DNA topoisomerase IIα promoter (Kotecha *et al.* 2008).

The collaboration was established in order to further characterise the effects of GWL-78 on alternative biological systems.

6.2.1 GWL-78 Interaction with DNA

GWL-78 was provided in powdered form and, under the direction of Mr Irfaan Dawood, suspended in 50 % v/v methanol to the desired molar concentration and used to prepare the drug-DNA adducts (2.5.5) at a 4:1 molar ratio (Figure 6.2). When visualised on an SDS PAGE gel using a Fujifilm FLA-5000 phosphoimager with a blue laser (473 nm) at 550 V and a green filter (532 nm), the drug-DNA adduct gave an incomplete shift when compared to the DNA only control. Furthermore, the partial shift was not eliminated when the samples were run on a native PAGE gel, implying that the shift was not induced by SDS, and a DNA-methanol incubation following the adduct preparation protocol failed to produce any shift, eliminating methanol as a potential affecter.

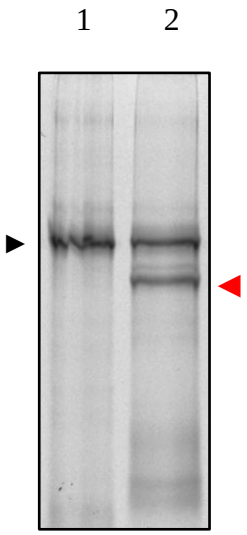


Figure 6.2 **GWL-78 Incubated with the Fork 1 Substrate.** The 15 % SDS PAGE gel shows the effect of GWL-78 binding on the Fork 1 substrate (substrate 10, Figure 2.5). Lane 1 shows the Fork 1 substrate prior to incubation with GWL-78, highlighted by the black arrow, while lane 2 shows the Fork 1 substrate after incubation GWL-78 at a 4:1 drug to DNA ratio; the resulting band shift is highlighted by the red arrow.

It was thought likely to be the result of unsaturated binding of GWL-78 to the DNA substrate. Therefore, electrophoretic mobility shift assay (EMSA) experiments were performed by titrating GWL-78 to the Fork 1 substrate from a 0 to a 32:1 molar ratio (Figure 6.3).

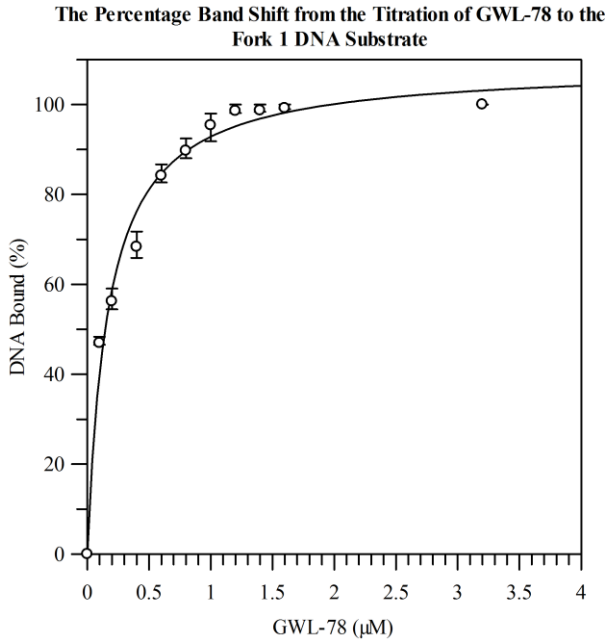


Figure 6.3 **Drug-DNA Adduct of GWL-78 with the Fork 1 Substrate.** The graph shows the percentage band shift by GWL-78 binding to the Fork 1 substrate (substrate 10, Figure 2.5) at increasing drug to DNA molar ratios. The lowest concentration showing a 100 % shift of the 100 nM Fork 1 substrate is 1.2 μM GWL-78, with a corresponding K_d of 0.1694 ± 0.0214 μM. Data is fit to a single site ligand binding curve.

This showed a total shift, indicating that GWL-78 binding to the DNA substrate was saturated at 12:1 drug to substrate ratio, and a K_d of 0.1694 ± 0.0214 μM. Curiously the observed change was a suprashift rather than a supershift, which is typically anticipated when studying DNA binding interactions (Fried *et al.* 1981; Garner *et al.* 1981). However, the downward shift may have been the result of a more compact shape being induced. This led to the theory that the free ssDNA ends of the Fork 1 substrate may have been being structured by GWL-78. To test the theory, a complimentary DNA strand to the X12-4 oligonucleotide was used to create a dsDNA 50mer equivalent of the Fork 1 substrate and run alongside the drug-DNA adduct.

Based on the theory detailed above, the dsDNA 50mer showed an even more compact shape than that of the Fork 1 substrate drug-DNA adduct. If the shift was indeed due to a more compact shape being induced by GWL-78, the drug-DNA adduct would have to have been more compact than the Fork 1 substrate, but less compact than the dsDNA 50mer. This led to the idea that maybe GWL-78 was stabilising the free ssDNA ends of the Fork 1 substrate by opportunistically binding when they formed weak interactions between one another, supported by the observed requirement for a large excess of GWL-78 to DNA. This would result in a more bubble-like substrate, similar to that described for the Fork 2 substrate (Table 4.2). The drug-DNA adduct of the Fork 1 substrate was run alongside the Fork 2 substrate and the dsDNA 50mer (Figure 6.4).

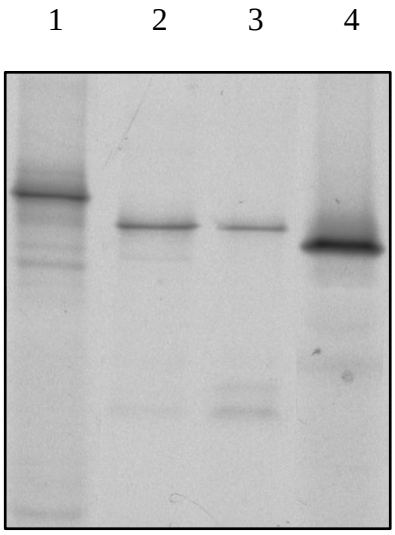


Figure 6.4 **Comparison of the Apparent Size Difference Resulting from GWL-78 Adduct Formation.** The 15 % SDS PAGE gel shows the Fork 1 substrate (substrate 10, Figure 2.5) and its GWL-78 adduct, the Fork 2 Substrate (substrate 11, Figure 2.5) and the complimentary dsDNA 50mer (substrate 36, Figure 2.5). Lane 1 shows the Fork 1 substrate prior to incubation with GWL-78, lane 2 shows the Fork 1 substrate after incubation with GWL-78 at a 16:1 drug to DNA ratio, lane 3 shows the Fork 2 substrate, and lane 4 shows the full complement Fork 1 dsDNA 50mer.

Based on these observations, it seemed that GWL-78 was ‘clamping’ the two free ssDNA ends together, therefore drug-DNA adducts were created for substrates containing one or two free ssDNA ends in different configurations (Figure 6.5). This suggested that the GWL-78 was clamping the free ssDNA ends within the 3’ Flap and Fork 1 substrates.

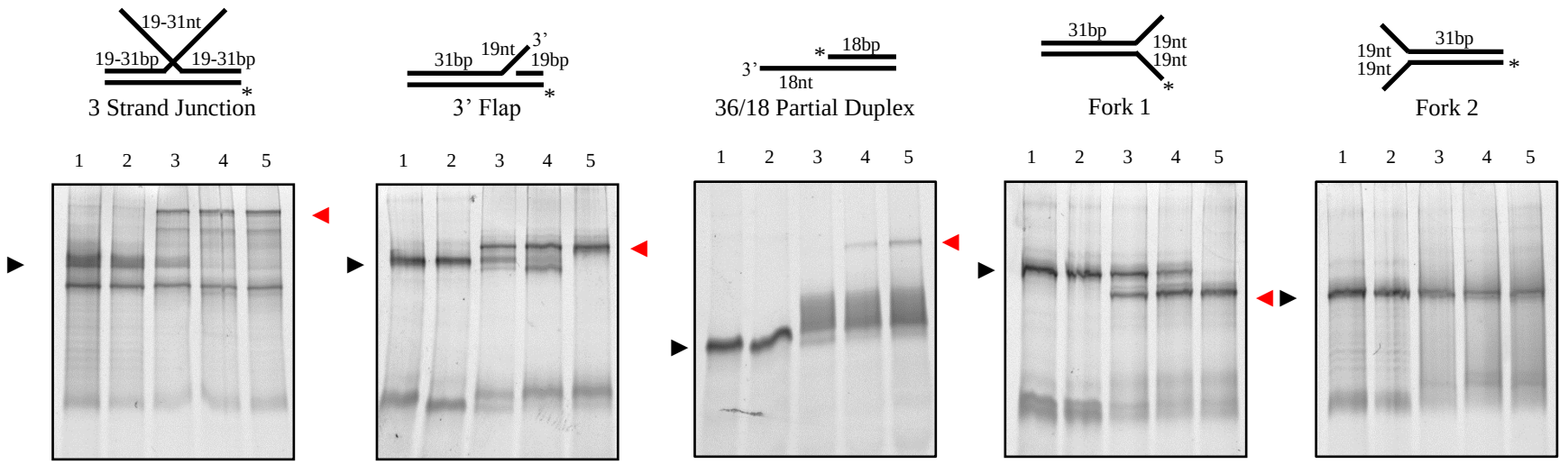


Figure 6.5 **Effect of GWL-78 on Substrates of Various Topologies with Free ssDNA Ends.** The 15 % SDS PAGE gels show the substrate incubated at varying molar ratios of DNA to GWL-78. Each gel lane is numbered 1 to 5; lane 1 shows the substrate without being incubated with GWL-78, while lanes 2 to 5 contain the substrate after being incubated with GWL-78 at 2, 4, 8, and 16:1 GWL-78 to DNA molar ratios respectively. Above each gel is a representative diagram of the substrate relating to that gel, substrates 2, 3, 7, 10, and 11 respectively (Figure 2.5). The solid black triangle identifies the corresponding band for each of the primary substrates being screened against, while any red triangles highlight any shift associated with incubation of GWL-78.

Moreover, in the case of the 3 Strand Junction and 50/19 Partial Duplex substrates, it was capable of clamping the ssDNA end of one molecule to the ssDNA end of another.

A final experiment was performed to observe the ssDNA clamping potential of GWL-78, where it was incubated with the constituent strands of the Fork 1 substrate to observe whether it could stimulate annealing. This was performed by incubating the constituent oligonucleotides of the Fork 1 substrate in annealing buffer. However, instead of heating and cooling slowly, as described under the substrate annealing section of chapter two, the mixture was used directly for drug adduct preparation with increasing molar ratios of GWL-78 (Figure 6.6). Although GWL-78 did not support the total annealing of the ssDNA molecules, there was a reproducible correlation between the amount of GWL-78 and the amount of dsDNA formed.

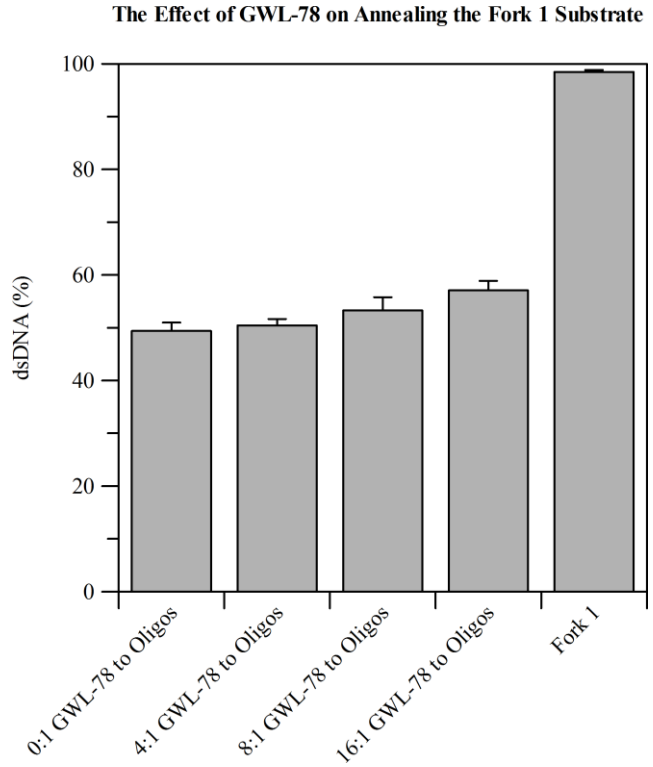


Figure 6.6 **DNA Annealing by GWL-78.** The graph shows the percentage of Fork 1 substrate (substrate 10, Figure 2.5) formed at increasing drug to DNA molar ratios. The leftmost set of data shows the amount of dsDNA formed when the Fork 1 constituent oligonucleotides are incubated without the presence of GWL-78. The second, third and fourth sets of data show the amount of dsDNA formed in the presence of GWL-78 at 4, 8, and 16:1 GWL-78 to DNA molar ratios. Finally, the rightmost set of data shows the amount of dsDNA formed under the standard substrate annealing conditions.

6.2.2 The Effect of GWL-78 on EcoRI24I Translocase Activity

Before carrying out experiments determining the effect of GWL-78 on the helicase activities of PFEIF4A or PFDH60, experiments on the type IC restriction modification (RM) enzyme EcoR124I were performed. EcoR124I is made up of a combination of three subunits; HsdR, HsdM, and HsdS. These subunits form complexes referred to as methylase (Mtase), R1 and R2. The Mtase is made up from two HsdM and one HsdS, abbreviated as M2S1. The R1 complex consists of one HsdR with the Mtase (R1M2S1) and the R2 complex consists of two HsdR subunits with the Mtase (R2M2S1). Only the R2 complex is restriction-proficient, however it is also less stable than the R1 and the Mtase complexes (Janscak *et al.* 1998).

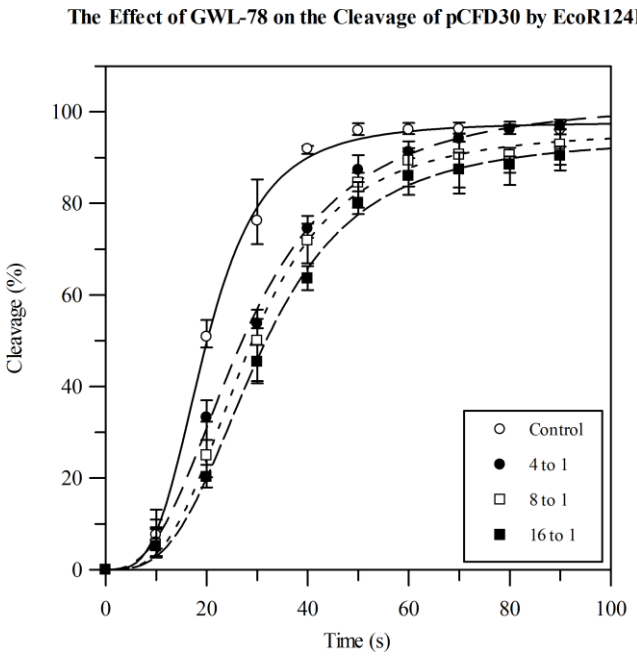


Figure 6.7 **Effect of GWL-78 on EcoR124I Cleavage.** The graph shows the percentage of pCFD30 cleaved by the EcoR124I R2 complex in the presence of increasing concentrations of GWL-78. The white circles represent the percentage of PCFD30 at the given time-points, and the solid line represents their line of best fit. The black circles represent the percentage of the 4:1 GWL-78 pCFD30 adduct at the given time-points, and the wide-spaced dashed line represents their line of best fit. The white squares represent the percentage of the 8:1 GWL-78 pCFD30 adduct at the given time-points, and the wide-spaced dotted line represents their line of best fit. Finally, the black squares represent the percentage of the 16:1 GWL-78 pCFD30 adduct at the given time-points, and the close-spaced dashed line represents their line of best fit. Data is fit to a cooperative ligand binding curve.

While EcoR124I was chosen as it is a relatively well understood enzyme, it had been extensively studied locally, and therefore clones and purification strategies were readily available from Dr James Youell (University of Portsmouth, Hampshire, UK).

90 second time-course experiments with 10 second time-points were carried out under the conditions described under the EcoR124I cleavage assay (2.3.8), and in the presence of pCFD30 GWL-78 adduct, at molar ratios of 4, 8, and 16:1 drug to DNA (Figure 6.7). Repeating the experiment without prior incubation of GWL-78 with pCFD30 at the same molar ratios had no notable effect on the rate of cleavage.

These experiments showed that EcoR124I was slower to cleave the available pCFD30 substrate, possibly due to GWL-78 slowing translocation which became more noticeable as the concentration of GWL-78 increased.

6.2.3 The Effect of GWL-78 on PFEIF4A Helicase Activity

Experiments were performed to see the effects of GWL-78 on the helicase activity of PFEIF4A. Based on the observations described in 6.2.1, GWL-78 Fork 1 adducts were created by incubating GWL-78 with the Fork 1 substrate at a 16:1 molar ratio. This was used in helicase 30 minute time-course experiments, with 20 µL 5 minute time-points. Each time-point reaction was carried out as described for the helicase assay with a 50:1 helicase to substrate ratio (5:0.1 µM), at 37 °C and stopped with 5 µL of 5 × termination buffer (Figure 6.8).

These experiments showed no discernible difference between PFEIF4A unwinding of the Fork 1 substrate and the GWL-78 Fork 1 adduct. The same experiment was carried out using the BOXTO helicase assay due to its increased resolution of the intermediate unwinding states (Figure 6.9).

These experiments suggest that although GWL-78 does not totally inhibit unwinding, it does slow down the initial rate of PFEIF4A helicase activity on the Fork 1 substrate by approximately 30 %, down from 185.00 ± 8.40 to 133.30 ± 3.95 fmol min^{-1}.

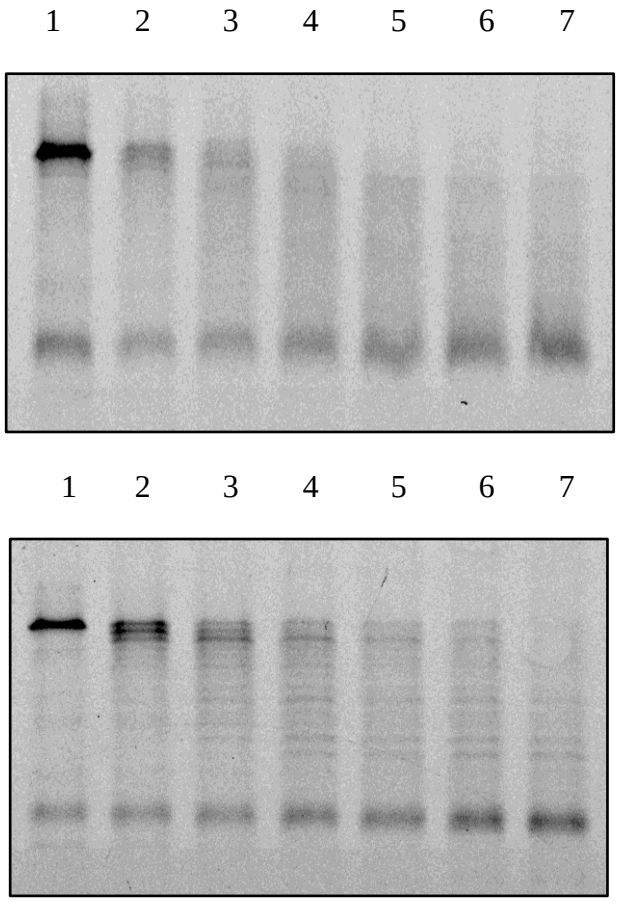


Figure 6.8 **Effect of GWL-78 on PFEIF4A Helicase Activity.** The 15 % SDS PAGE gels are examples of the time-course experiments of PFEIF4A unwinding the Fork 1 substrate (substrate 10, Figure 2.5) (top) and the GWL-78 Fork 1 adduct (bottom). The lanes in both gels are labelled 1 to 7. Lane 1 shows the 0 minute time-point and lanes 2 to 7 show the 5, 10, 15, 20, 25, and 30 minute time-points.

6.3 Aptamers

Protein specific targeting by drugs is lengthy as it relies on basic biological knowledge of the target protein (Palchaudhuri *et al.* 2007) and so aptamers have emerged as an alternative strategy. The ability to raise specific aptamers against target molecules has led to their proposed use in therapeutics and as diagnostic tools. As therapeutics, aptamers could be used as inhibitors, similar to pharmaceutically developed drugs.

As a diagnostic tool, they could be used as biological markers, as their high specificity and affinity lends potential as an alternative to monoclonal antibodies (Stoltenburg *et al.* 2007). They have the advantage of being chemically synthesised easily (Tuerk *et al.* 1990) and

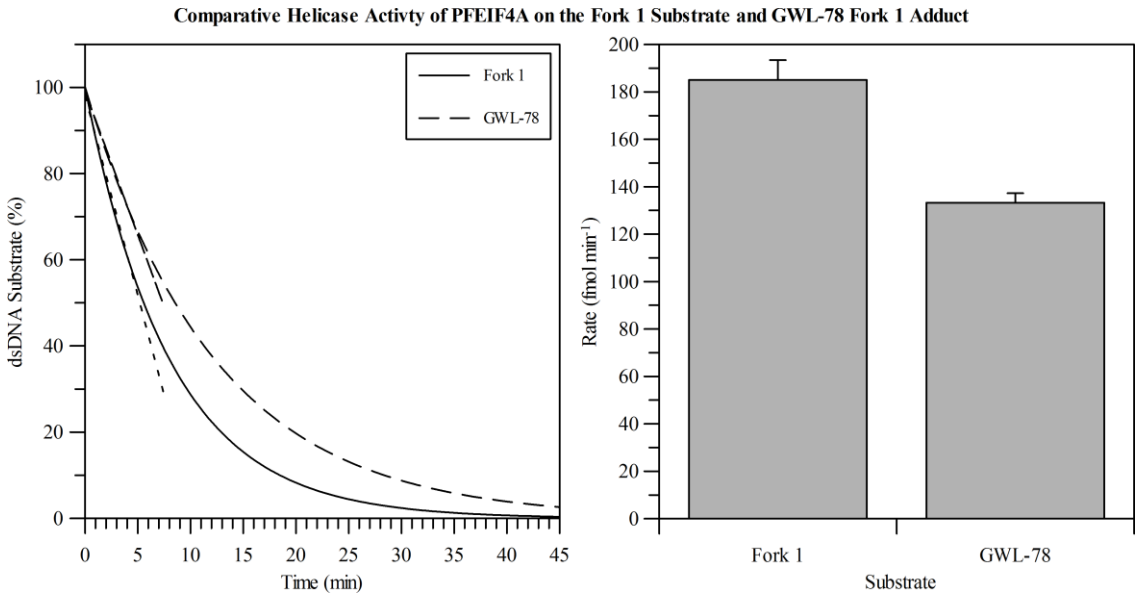


Figure 6.9 **BOXTO Helicase Assay of PFEIF4A Against the GWL-78 Fork 1 Adduct.** Graph A shows the lines of best fit for the averaged results from triplicate experiments on PFEIF4A unwinding activity on the Fork 1 substrate (substrate 10, Figure 2.5), the solid line, and the GWL-78 Fork 1 adduct, the wide-spaced dashed line. Data is fit to a single site exponential decay curve .The dotted line and the close-spaced dashed line are the lines of best fit used to calculate the initial rates of unwinding (first 5 minutes) of the Fork 1 substrate and the GWL-78 Fork 1 adduct respectively. The error bars have been omitted for clarity. Graph B shows the difference between the initial rate of unwinding (fmol min^{-1}) by PFEIF4A on the Fork 1 substrate and the GWL-78 Fork 1 adduct.

unlike antibody production, do not require a relatively costly host organism, and can be made chemically stable as not to trigger an immune response.

Furthermore, as biosensors they could be further developed to discriminate between species, for example *P. falciparum* proteins within human blood samples, helping to develop specific medical treatments. If linked with structural studies, they could also prove a quick way to find critical binding sites which could then be used to raise drugs against.

6.3.1 The Effect of HsdR and Mtase Aptamers on EcoRI24I Translocase Activity

5’ Biotinylated aptamers, generated by Dr Daniel Fordham (University of Portsmouth, Hampshire, UK) against the HsdR subunit and Mtase complex (Table 6.1), were used as

Name	Sequence
HBA1	5′-CCAAGCGGGCGCGAGAACAACGTACGGT-3′
HBA3	5′-CCGAGCAGAAGGTGCGATGCTTTATGGT-3′
HBA8	5′-CCGAGCAGAAGGTGCGATGCTTTATGGT-3′
HBA14	5′-CCCGGGCGGACAGACGGGTGTTCGTTGC-3′
HBA15	5′-CCCGGGCGGACAGCAGGCTGTTCGTTGC-3′
HBA16	5′-CCCGTAGCGTACAGTCCCAGCTTGGGAT-3′
HBA20	5′-CCACAACGAGCCGGGATTGCGGATTGCA-3′
MBA1	5′-CCTGCTGATTCGTTGGGTGTTGTGTGGT-3′
MBA2	5′-CCCGTANCGTACAGTCCCAGCTTGGTAT-3′
MBA3	5′-CCCTGAGTCATAGTGTTCACGTGTT-3′
MBA4	5′-CCACAACGAGCCGGGATTGCGGATTGCA-3′
MBA5	5′-CCCTGAGTCATAGTGTTCACGTGTT-3′
MBA6	5′-GGCGAATTGGGCCCCGAGGGGTCGACT-3′
MBA7	5′-CCCGTANCGTACAGTCCCAGCTTGGTAT-3′
MBA8	5′-GGCGAATTGGGCCCGAGTCGCATGCTCCC-3′
MBA9	5′-CCGGTTGGGATGTGNCNCCCCNNATANN-3′

Table 6.1 **HsdR and Mtase Binding Aptamers.** The table shows each aptamer sequence with its name as assigned by Dr Daniel Fordham; HsdR binding aptamers (HBAs) and Mtase binding aptamers (MBAs). MBAs 2, 7, and 9 all contain non-determined nucleotides (N).

primary probes to dot blot against various concentrations of their respective targets, and developed using a monoclonal Anti-Biotin Peroxidise Conjugate (Abcam) (Figure 6.10).

The dot blots were scrutinised using densitometry (Schneider *et al.* 2012) and the two aptamers with the highest density were used for EcoR124I cleavage assays. In short, 90 second time-course experiments with 10 second time-points were carried out in the presence of 10 µM HBA6, HBA15, MBA6, or MBA8 (Figure 6.11).

These experiments showed that although the initial binding of the EcoR124I machinery was not affected by the presence of any of the aptamers, the subsequent translocation leading to cleavage was greatly inhibited in the presence of MBA6 and to a lesser extent MBA8.

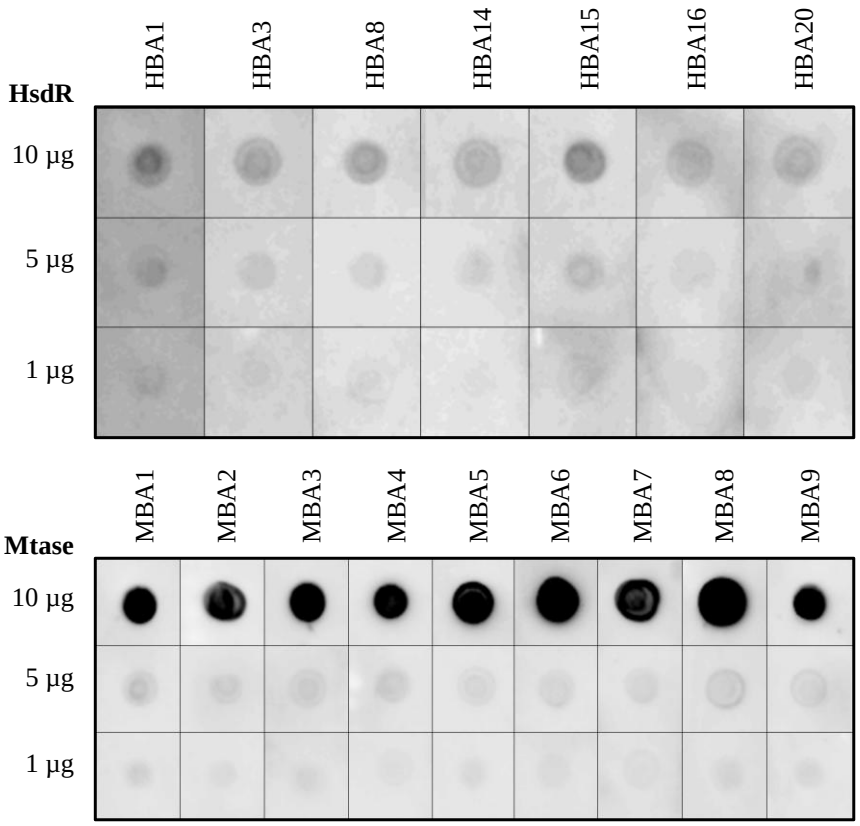


Figure 6.10 **HBA and MBA Dot Blots against the EcoR124I HsdR subunit and Mtase Complex.** Both dot blots have the aptamers blotted across the *x*-axis and the HsdR (top) or Mtase (bottom) blotted in decreasing concentration down the *y*-axis.

6.3.2 Generating Aptamers against PFEIF4A

Successive rounds of SELEX and PCR amplifications were performed on 100 µg His_6-T-tagged PFEIF4A immobilised to a Ni-NTA centrifuge column. PCR amplified SELEX products were run on a native PAGE gel and post-stained with 1 mg mL^{-1} ethidium bromide in 1× TAE buffer (Figure 6.12).

Although a clear product band could not be seen after multiple successive SELEX rounds, cloning of the products was continued due to similar previous observations which resulted in successful cloning. However, the few recombinant colonies that were obtained did not yield any aptamer sequences upon sequencing.

In an attempt to increase the yield of PCR product, 1 µL of a 1000-fold dilution of the final round of SELEX was used as the template for PCR amplification. The rationale being that

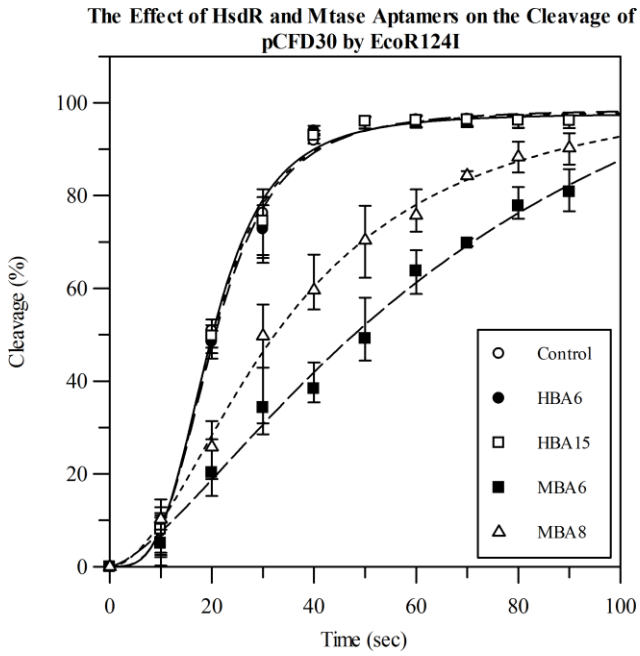


Figure 6.11 **The Effect of HsdR and Mtase Aptamers on EcoR124I Cleavage.** The graph shows the percentage of pCFD30 cleaved by the EcoR124I R2 complex in the presence of HsdR and Mtase aptamers. The white circles represent the percentage of PCFD30 at the given time-points for the control experiment, and the solid line represents their line of best fit. The black circles represent the percentage of pCFD30 at the given time-points in the presence of HBA6, and the wide-spaced dashed line represents their line of best fit. The white squares represent the percentage of pCFD30 at the given time-points in the presence of HBA15, and the wide-spaced dotted line represents their line of best fit. The black squares represent the percentage of the pCFD30 at the given time-points in the presence of MBA6, and the close-spaced dashed line represents their line of best fit. Finally, the white triangles represent the percentage of pCFD30 at the given time-points in the presence of MBA8, and the close-spaced dotted line represents their line of best fit. Data is fit to a cooperative ligand binding curve.

having a more dilute template DNA sample, as opposed to a whole SELEX round product, would increase the likelihood of a cooperative encounter between the PCR machinery. The PCR product was run on a native PAGE gel (Figure 6.13). This product was then cloned using the pGEM-T Easy Vector system (Promega). Recombinant colonies were grown for DNA extraction and sequence quality purification using centrifuge columns (Machery-Nagel). The extracted DNA was then sequenced and their shapes predicted using the mfold and RNAfold web servers (Zuker 2003; Gruber *et al.* 2008) (Table 6.2). Although RNAfold is primarily designed for predicting RNA folding, observations made by Chen *et al.* (2012) have shown that short ssDNA is more flexible than short ssRNA. Therefore using RNAfold is valid in proposing an alternative shape to that of mfold. Many of the

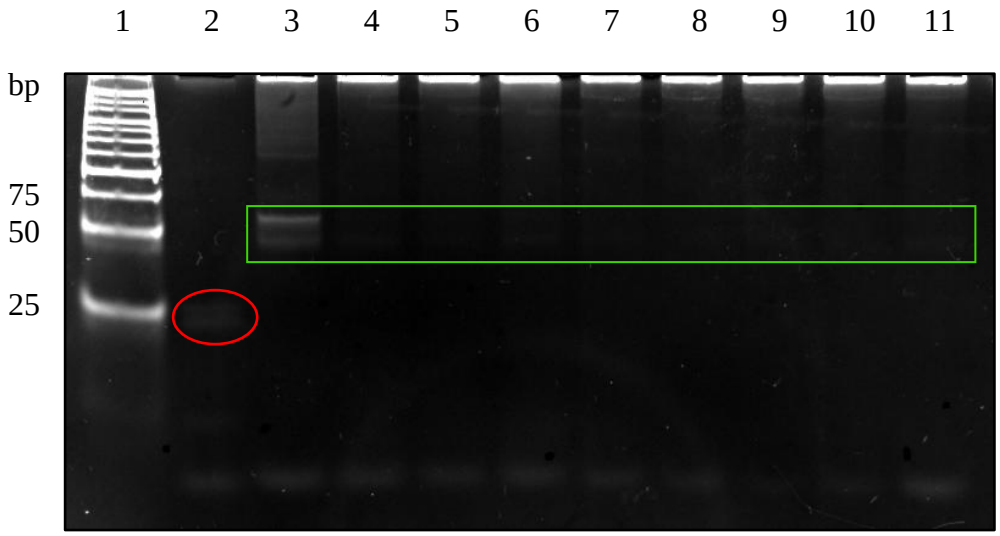


Figure 6.12 **Successive Round of SELEX to PFEIF4A.** The 15 % native PAGE gel shows the HyperLadder V marker (Bioline) in lane 1. Lane 2 shows the pre-screened library, while lanes 3 to 11 show the PCR amplified products from the first to the eighth round of SELEX respectively. The pre-screened library shows a faint band corresponding in size to approximately 25 bp (red), which is expected for the single-stranded 59 nt random library pool. Lane 3 shows numerous bands corresponding in size to approximately 50 bp (green), with the most predominant band slightly higher (~60 bp). Again, this is expected for the PCR amplified product as the dsDNA product (2 × 59 nt) was not denatured prior to loading. The dsDNA bands become fainter with each subsequent round of SELEX.

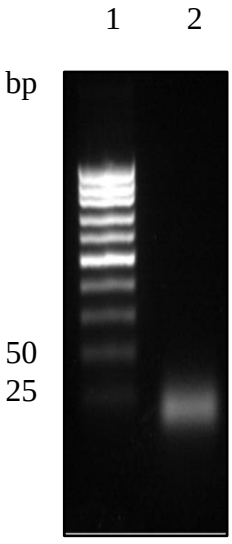


Figure 6.13 **PCR Amplification of Diluted Eighth SELEX Round against PFEIF4A.** The 15 % native PAGE gel shows the HyperLadder V marker (Bioline) in lane 1. Lane 2 shows the denatured PCR (59 nt) product obtained from the 1000-fold dilution of the eighth round of SELEX against PFEIF4A, with a band corresponding in size to approximately 25 bp.

Name	Sequence
4ABA1	5′-CCCACAGACATCAGGACGCTACATGGGGT-3′
4ABA2	5′-GGGCCAGGTGGTGTGGTGTTGCTGGGT-3′
4ABA3	5′-CCCACAGACATTAGGACCCTACGTGTGGT-3′
4ABA4	5′-CCCACAGACATTAGGACGCTACATGGGGT-3'

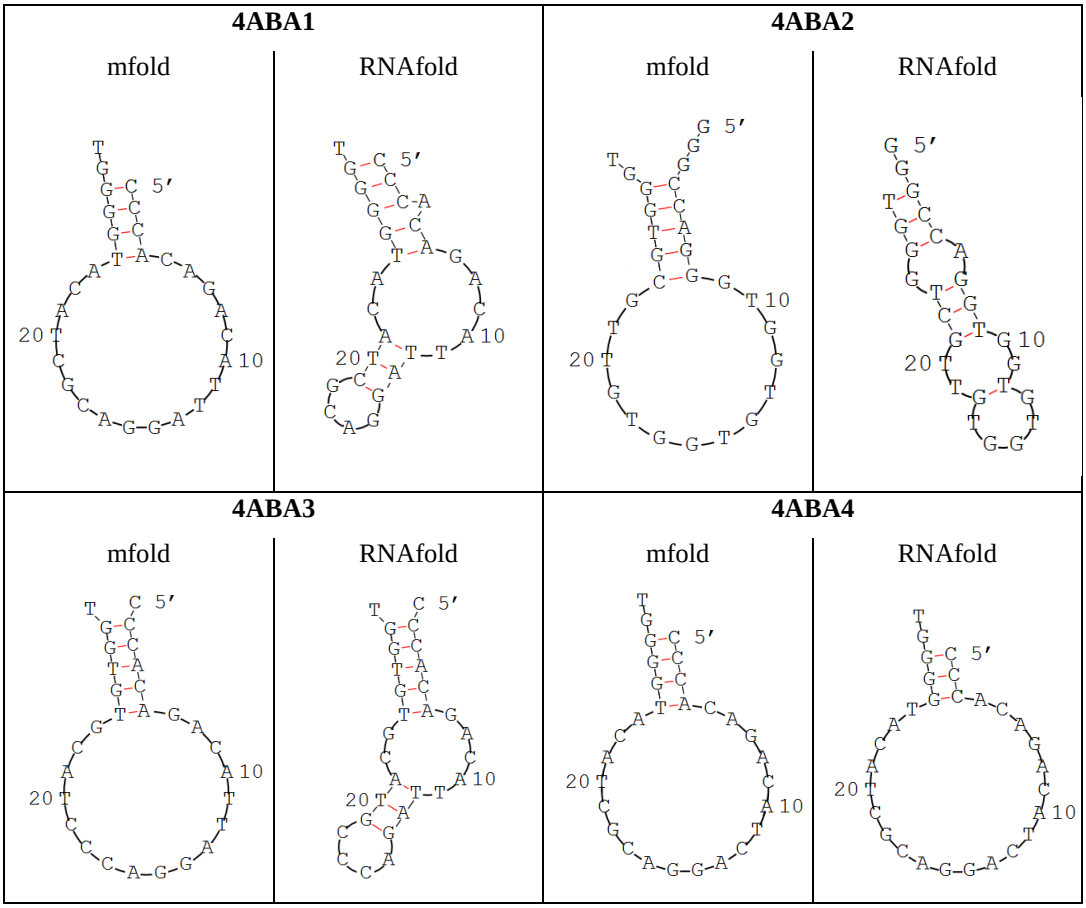


Table 6.2 **PFEIF4A Binding Aptamers.** The top table shows each aptamer sequence with its name; PFEIF4A binding aptamer (4ABA). The bottom table shows the predicted shapes of each of the aptamers based on the mfold (Zuker 2003) and the RNAfold web servers (Gruber *et al.* 2008).

sequences returned were multiple copies of the same sequence. 4ABA1 itself was a quadruplicate; this is likely due to the use of a relatively small number of template DNA molecules, resulting in the yield of large amounts of comparatively few aptamer sequences.

Initial dot blots performed as described for the EcoR124I aptamers could not be detected. As a result, dot blots were performed using alternative probing approaches, including pre-

incubating each aptamer with PFEIF4A at room temperature for 15 minutes prior to being immobilised (Figure 6.14). These experiments suggested that none of the aptamers could bind immobilised His_6-T-PFEIF4A and be detected above that of the immobilised aptamer alone.

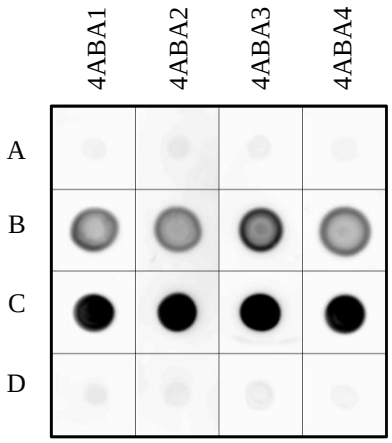


Figure 6.14 **4ABA Dot Blots against the PFEIF4A.** The aptamers 4ABA1 to 4 are blotted across the *x*-axis. The *y*-axis is labelled A to D. Row A is 2.5 µg of immobilised PFEIF4A only and row B is 2.5 µg of PFEIF4A pre-incubated with 1 µL of 100 µM aptamers for 15 minutes at room temperature prior to the mixture being immobilised. Row C is 1 µL of 100 µM immobilised aptamers only, and row D is 2.5 µg of immobilised PFEIF4A detected with each aptamer based on the standard dot blotting method.

6.4 Discussion

Due to the threat posed to some of the world's most vulnerable people and the lack of a long-term solution, compounded by wide-spread resistance to vector control strategies and parasitic interventions (EANMAT 2003; Plowe *et al.* 2004), the pursuit of effective anti-malarial strategies remains a constant focus for the WHO (WHO 2008; WHO 2009; WHO 2010).

This chapter has described two strategies in targeting the PFEIF4A helicases as proposed by Tuteja (2007). They are the effect of the novel DNA-interactive PDB conjugate GWL-78, and raising aptamers to specifically target PFEIF4A.

6.4.1 GWL-78

Previous studies identified that GWL-78 binds dsDNA (Wells *et al.* 2006; Rahman *et al.* 2011). This work has presented the hypothesis that it is also capable of binding single-stranded DNA regions to one another. To explore this further, computational models of the Fork 1 substrate with and without GWL-78 were rendered using the Assisted Model Building with Energy Refinement (AMBER) package by Mr Irfaan Dawood (Figure 6.15). This model supports the theory that GWL-78 was 'clamping' the free ssDNA ends creating a more compact molecule.

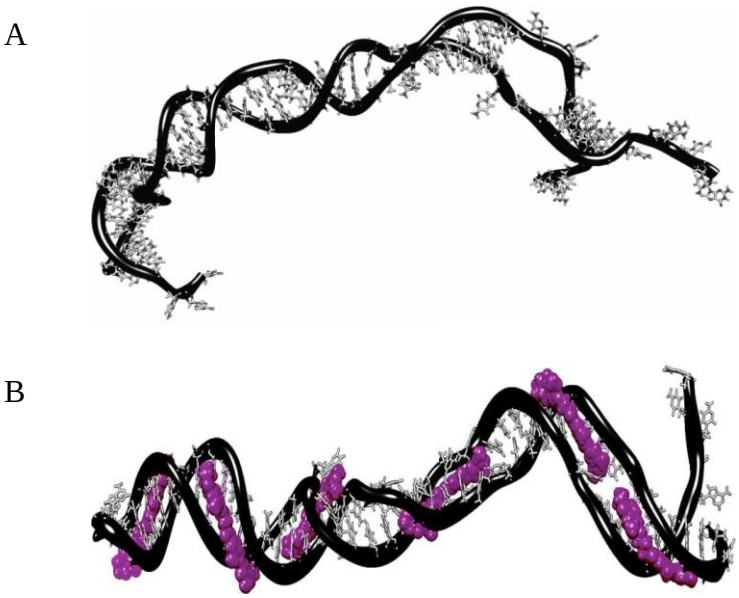


Figure 6.15 **AMBER Generated Models of the Fork 1 Substrate and its GWL-78 Adduct.** Schematic A shows the structure of the Fork 1 substrate (substrate 10, Figure 2.5) with the phophodiester backbone shown in black and the bases shown as stick models. Schematic B shows the Fork 1 substrate with six GWL-78 molecules (magenta) bound as described by Rahman *et al.* (2011). Note the forked regions of the substrate (rightmost) are seemingly 'clamped' into a duplex by the bound GWL-78.

In addition, the experiments performed in this chapter show that although GWL-78 does not inhibit either activity totally, it does have an effect on both EcoR124I translocase and PFEIF4A helicase activities.

6.4.2 Aptamers

Since being identified approximately twenty years ago (Ellington *et al.* 1990; Tuerk *et al.* 1990), there have been many studies recognising the potential of aptamers. This has led to the emergence of two distinct aptamer technologies, therapeutic and diagnostic. Therapeutic aptamers are generally used as inhibitors, and at least one commercially available therapeutic aptamer, Macugen (OSI Pharmaceuticals), has been approved in the United States of America (U.S.A.) (Doggrell 2005). Likewise, NeoVentures Biotechnology Inc. now offers a commercial aptamer-based diagnostic platform, which target molecules in a similar way to antibody technologies (Jayasena 1999). The approved use and success of these aptamer technologies is encouraging for their continued and expanding use for a wide range of applications and treatments of severe conditions and diseases.

The experiments using HsdR and Mtase binding aptamers showed potential for both diagnostic and therapeutic applications. All the aptamers seemed to show specific binding to the HsdR subunit. Due to time restraints only two aptamers for each target were used for further preliminary experiments. Neither HBA6 nor 15 had an effect on EcoR124I cleavage. However, MBA6 and 8 both showed inhibitory effects on EcoR124I cleavage. It may be that the aptamers interrupted the formation of the EcoR124I complex. Further gel shift experiments using the EMSA assay as detailed in chapter two could have been performed to test this theory.

SELEX performed against PFEIF4A only produced four aptamers. None of these showed significant detection of immobilised PFEIF4A. Time permitting, alternative techniques could have been explored, as the immobilisation of PFEIF4A to the nitrocellulose membrane may have affected its structure, distorting the potential aptamer binding sites. Therefore, a solution based detection strategy may have been more suitable. This could have been done by immobilising the His_6-T-PFEIF4A to a Ni-lined well and washing the bound PFEIF4A with the aptamers, followed by washing with the 2° HRP conjugated probe prior to detection.

CHAPTER SEVEN: CONCLUSIONS

7.1 Background

Helicases are a subclass of translocases that are capable of unwinding the nucleic acid duplexes. As a consequence they are vital to all living organisms, playing key roles in all cellular processes, such as DNA repair and replication (Cobb *et al.* 2002), transcription (Ciampi 2006), translation (Rogers *et al.* 1999), recombination (Singleton *et al.* 2004), RNA maturation and splicing (Staley *et al.* 1998), RNA degradation (Py *et al.* 1996), ribosome biogenesis (Venema *et al.* 1995), nuclear export processes (Kanai *et al.* 2004), and mitochondrial genome replication (Korhonen *et al.* 2003).

Plasmodium falciparum is responsible for over 90 % of worldwide cases of malaria in humans, with the highest reported complications and mortality rates when compared to all other human malaria-causing *Plasmodium* (WHO 2009; WHO 2010). Interest in the control of malaria has become increasingly focused on the worsening situation in sub-Saharan Africa and its contributory factors (Greenwood *et al.* 2002) such as climate instability and global warming (Hay *et al.* 2002; Tanser *et al.* 2003), healthcare breakdown due to war and civil-unrest (Bloland *et al.* 2002), interaction with and increased susceptibility to malaria due to HIV infection (Whitworth *et al.* 2000; French *et al.* 2001; Mount *et al.* 2004; Kublin *et al.* 2005), and the increased rate of travel between non-endemic and endemic areas. However, drug-resistant parasites and insecticide-resistant *Anopheles gambiae* mosquitoes are thought to be the likely main cause (Snow *et al.* 2001; Korenromp *et al.* 2003).

7.2 Observations made in this Study

The *P. falciparum* genome has been fully sequenced (Gardner *et al.* 2002), and several studies exploring the exploitation of the parasites' proteins as anti-malarial targets have been performed. The helicases described within this study, PFEIF4A and PFDH60, were first proposed as potential drug-targets by Dr Renu Tuteja (Malaria Group, International

Centre for Genetic Engineering and Biotechnology, New Deli, India) (Tuteja 2007). The work described here was aimed at expanding the understanding of PFEIF4A and PFDH60 through biochemical and structural analyses, and the high-throughput generation and effects of novel inhibitors. It should be noted that the PFDH60 protein discussed within this study is the full-length gene product (87 kDa). It is referred to as its 60 kDa namesake for the benefit of continuity regarding previous studies (Pradhan *et al.* 2005; Pradhan *et al.* 2006; Pradhan *et al.* 2007), however the work described in this study is entirely novel.

This was performed firstly by optimising the growth and purification conditions as the original strategies (Pradhan *et al.* 2005; Pradhan *et al.* 2007) proved inadequate for the yields required for full biochemical and structural analysis. Initial biochemistry reported PFEIF4A helicase and ATPase activities on a limited range of substrates (Pradhan *et al.* 2008). This study has expanded to use numerous substrates with diverse topologies. It has identified an absolute requirement for a single-stranded overhang and a 3'-5' polarity bias by PFEIF4A. It has detailed the effects of different DNA topologies on its ATPase activity, its preference for dsDNA, and DNA:RNA and RNA:DNA hybrid duplexes over dsRNA. It has also explored the effects of different divalent metal ions, its optimum temperature for unwinding activity, and its NTP usage. Dynamic Light Scattering (DLS) experiments have expanded on the observations made while examining the optimum working temperature of PFEIF4A to show that it becomes structurally unstable with temperature increases above its 37 °C optimum and that its unwinding activity is susceptible to inhibition by DNA binding compounds.

These observations have led to two questions. Firstly, is the effect of temperature significant? Secondly, can PFEIF4A and PFDH60 be exploited as anti-malarial targets?

7.2.1 Is the Effect of Temperature Significant?

Both proteins were liable to degradation and loss of activity during prolonged storage and incubation at increasing temperatures. PFDH60 was the less stable of the two proteins, displaying a total loss of activity within one week regardless of storage buffer or temperature. PFEIF4A was stable during storage at -20 °C in 10 % glycerol, but degraded at higher temperatures (4 °C) or without glycerol after a single freeze-thaw.

This observation is not surprising as unwinding activity is tightly regulated by the cell which invests substantial effort in maintaining more stable dsDNA over less stable ssDNA. The fact that PFDH60 degrades over a short space of time and that its expression within the parasite is limited mainly to the schizont life stage (Pradhan *et al.* 2005), suggests it is required for a specific role during the massive growth phase observed prior to the release of thousands of merozoites back into the bloodstream. Therefore, it is only required to be expressed for a short period of time before being recycled through the cells protein degradation process. Its susceptibility to degradation is also supported by the PHYRE2 server predicted structure of PFDH60 showing large unstructured regions, which are potentially vulnerable to degradation by proteolytic enzymes.

As described earlier, storage of PFEIF4A was more successful than that of PFDH60. Based on its expression throughout the parasitic life cycle (Pradhan *et al.* 2007), its similarity with human EIF4A1, and the identification of potential binding partners (Tuteja *et al.* 2009), it is suggested that PFEIF4A may contribute to the orthologous *P. falciparum* EIF4 complex. If so, like the human EIF4 complex it would serve as the main regulatory element of protein expression (Conroy *et al.* 1990; Yoder-Hill *et al.* 1993), and so would be required to be ubiquitously expressed throughout the cells life-span.

Ironically Julius Wagner-Jauregg received the 1927 Nobel Prize in medicine for inoculating neurosyphilis patients with malaria to induce a fever to treat the *Treponema pallidum* infection. However, temperature has long been acknowledged to affect the growth and development of *P. falciparum* (Kwiatkowski 1989). Curiously, inhibition of parasite growth has been reported at temperatures as low as 40 °C, despite febrile temperatures associated with malarial infection often exceeding this. Furthermore, mimicking febrile temperatures in combination with the administration of common anti-malarial drugs shows marked inhibition of field isolated parasites (Long *et al.* 2001; Aunpad *et al.* 2009). These observations may lend credibility to the traditional hydrotherapy strategies for the treatment of malaria infections performed by African and Malaysian tribes (Jennings 2006). Then, 3 to 4 hours before the onset of each fever associated with the release and re-infection of red blood cells (RBCs) by the parasite, the sufferer is wrapped in a blanket and placed into the sun to raise their body temperature and induce sweating, at which point they are submerged into cold water. The sufferer is kept bound and cool in between the process being repeated during the course of the infection.

The dramatic decrease in unwinding activity by PFEIF4A at temperatures greater than 37 °C coupled with its decreasing structural integrity around the same temperature, based on the observations made during the DLS experimentation, may help to provide a molecular link to these observations. This could prove an interesting direction for research towards an effective combination treatment against malaria.

7.2.2 Can PFEIF4A and PFDH60 be exploited as Anti-Malarial Targets?

PFEIF4A has been shown to share a high degree of identity to the human EIF4A1 (66.7 %) and its helicase and ATPase activities are affected by various DNA-binding compounds (actinomycin, 4',6-diamidino-2- phenylindole (DAPI), daunorubicin, ethidium bromide, netropsin and noglamycin) (Pradhan *et al.* 2008). This study has also shown that the unwinding activity of PFEIF4A is affected by the novel DNA-interactive PDB conjugate GWL-78. Previous studies have shown that parasite growth was inhibited by anti-PFEIF4A antibodies, and with PFEIF4A dsDNA in RNA interference experiments (Pradhan *et al.* 2007).

Although PFEIF4A is relatively similar in sequence and potential structure to its human orthologue EIF4A1, the biochemical analysis undertaken during this study has elucidated substantial differences between the two proteins' activities. Drugs that target the protein directly could take advantage of the two protein NTP usages. Both proteins have an optimal unwinding activity in the presence of ATP. However, human EIF4A1 can only utilise dATP as an alternative hydrolysis source (Du *et al.* 2002). The Tuteja group (Pradhan *et al.* 2008) observed that PFEIF4A utilised dATP, dCTP, and UTP, whereas this study has also observed significant unwinding in the presence of dGTP. Another distinct difference is between the nucleic acid substrates each protein acts upon. Human EIF4A has only been observed to unwind dsRNA, as well as RNA:DNA and DNA:RNA hybrid substrates (Rogers *et al.* 2001; Rogers *et al.* 2002). Although previous studies have shown PFEIF4A to unwind simple short RNA substrates (Pradhan *et al.* 2007; Pradhan *et al.* 2008), this study has shown that PFEIF4A is incapable of unwinding a dsRNA forked substrate despite being capable of unwinding dsDNA, and DNA:RNA and RNA:DNA hybrid forked substrates. Furthermore, human EIF4A is capable of unwinding blunt-ended

substrates (Rogers *et al.* 2001; Rogers *et al.* 2002), whereas PFEIF4A has an absolute requirement for a single-stranded overhang to support unwinding.

If the differences between human EIF4A1 and PFEIF4A are directly related to their structures, for example the size of the ATP binding pocket, then specific drugs that bind to PFEIF4A but not human EIF4A1 binding sites could be used to inhibit the unwinding or ATPase activities (Spencer *et al.* 2011) without affecting the parasite hosts' EIF4A proteins.

This study has identified that PFDH60 has very little identity to any human proteins. Previous studies have identified that PFDH60 unwinding and ATPase activities are affected by various DNA-binding compounds (actinomycin D, daunorubicin, ethidium bromide, netropsin, and nogalamycin) and that parasitic growth is affected by anti-PFDH60 antibodies, and PFDH60 dsRNA (Pradhan *et al.* 2006). These observations support that firstly it is an effective target, but also one with potentially minimal side effects to the host system. Its schizont limited expression may also help to specifically target PFDH60, as treatments that target the parasitised red blood cell (pRBC) could be directly injected into the bloodstream.

7.3 Future Work

Due to time restraints and unforeseen complications while undertaking this study, the pursuit of some lines of experiments had to be neglected or cut short. If time permitted, investigation into the biochemical and structural characterisation of PFDH60 would have been performed to the same extent as PFEIF4A. Further experimentation into dsRNA unwinding would have been explored to see if PFEIF4A could unwind RNA duplexes of sufficiently low thermostability to expand on observations made in this study in support of those made by Pradhan *et al* (Pradhan *et al.* 2007; Pradhan *et al.* 2008).

The BOXTO based unwinding assay and luciferase ATPase assay would have would have made high-throughput analysis of both PFEIF4A and PFDH60 helicase activities relatively easier. Work had been started on developing a high-throughput SPR binding analysis method. This involved attaching an array of thiolated DNA substrates to a high index gold

prism (Horiba) using the Genetix QArray2 micro-array printer. This could be visualised using the Horiba SPRi-Lab+ system, which converts the resonance into an image (Figure 7.1). Unfortunately, PFEIF4A non-specifically bound to the gold surface despite passivation with bovine serum albumin (BSA). Concern over the amount of time it could take to explore more effective means of passivation, and other experiments requiring attention, resulted in the technique being abandoned.

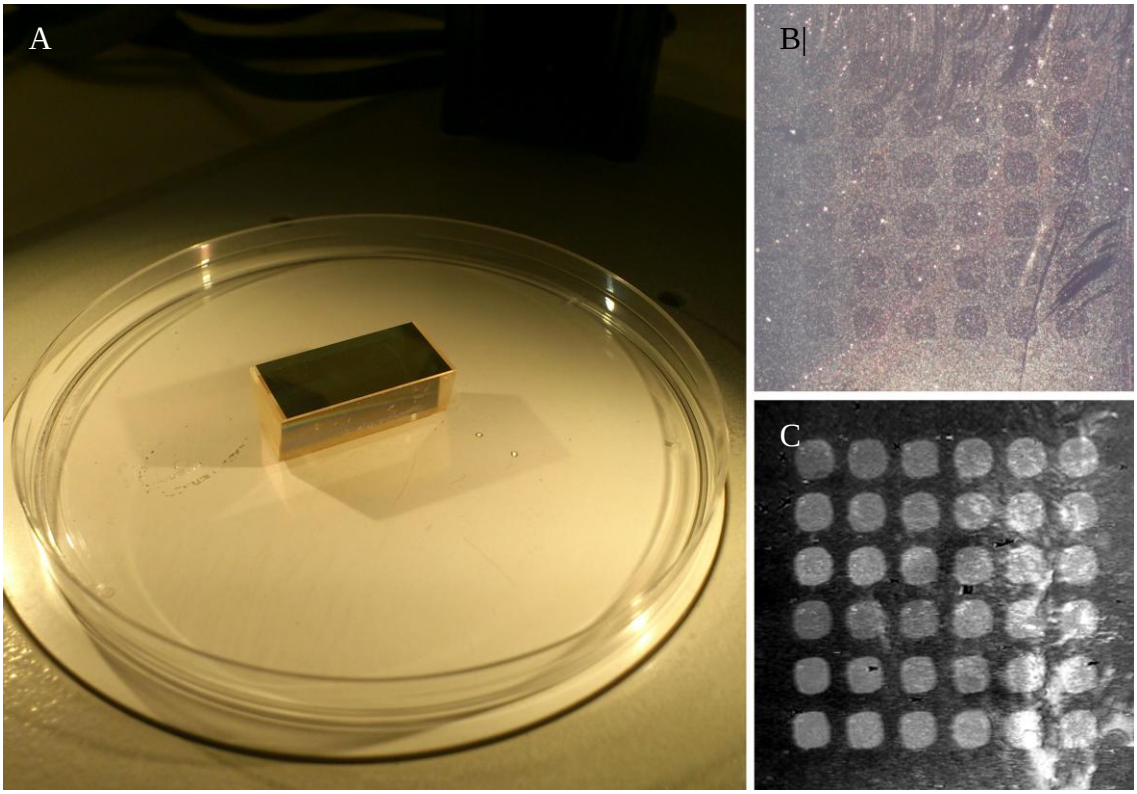


Figure 7.1 **SPRi-Lab+ DNA Array.** Panel A shows the high index glass prism and panel A shows the DNA substrates arrayed on its surface under a 1000 × magnification optical microscope. Panel C shows the image created using the SPRi-Lab+ system from the resonance profiles obtained using the arrayed prism. The 6 × 6 grid of dots correspond to three copies of the 12 Özsoy substrates; substrates 1 to 12 as described in Table 2.3, incorporating thiolated oligonucleotides 1, 4, 9, and 11 (Appendix 1).

The prevalence of single molecule experiments being published has greatly increased in recent years. There were 979 articles published in 2002, available through the PubMed website when searching for ‘single molecule’ in all fields. The number of articles published in 2011 was 2529, and there were already 2043 new publications in 2012 at the beginning of September. Due to the increased interest and the robust data generated from

single molecule studies, a collaboration was established with Prof. David Bensimon (Ecole Normale Superieure (ENS), Paris, France) to study PFEIF4A and PFDH60 using a magnetic tweezers system (PicoTwist). When PFEIF4A was first purified, but before the biochemistry was performed, experiments using the basic hairpin substrate were performed (Manosas *et al.* 2010). However, no data was generated. After performing the biochemistry described in this study and discovering that although PFEIF4A could unwind hairpin containing substrates, it also had an absolute requirement for a free single-stranded overhang in order to unwind a DNA duplex, a new substrate was designed (Figure 7.2). Unfortunately, the opportunity to perform any experiments with the novel substrate tailored to the specific requirements of PFEIF4A was not available due to the time restraints of this study.

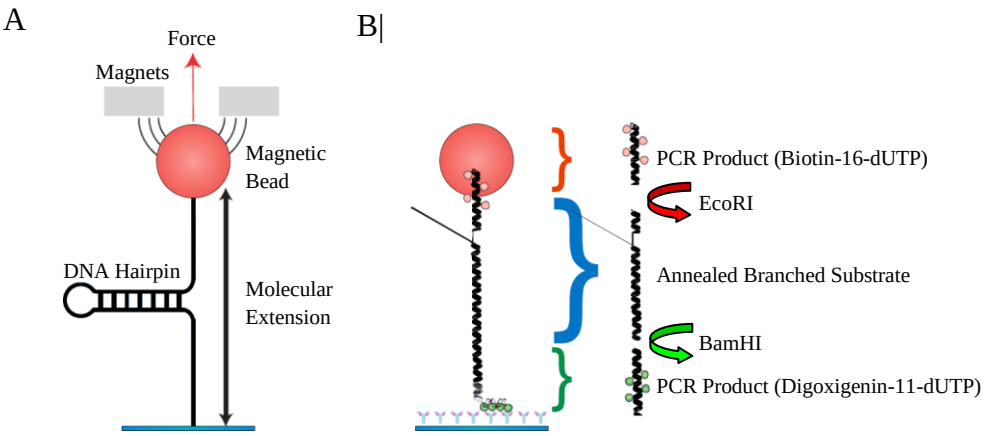


Figure 7.2 **Substrates for use in the Magnetic Tweezers Single Molecule System.** Panel A shows the typical hairpin substrate used by the Bensimon Group, adapted from Manosas *et al* (2010). Panel B shows the substrate specifically designed to observe PFEIF4A unwinding using the magnetic tweezers system, adapted from Firman *et al* (2012). Construction of the substrate involves the PCR amplification and gel purification of a digoxigenin and a biotin, incorporating fragments for surface and bead attachment respectively. The branched substrate is gel purified after annealing of its two component strands (Invitrogen). Each end fragment has either a BamHI or EcoRI restriction site for ligation to the annealed branch substrate. The final substrate is then gel purified.

The relative ease to generate and characterise aptamers has already been described in this study, and limited experimentation on aptamers targeting PFEIF4A has been undertaken. The effect of aptamers on cleavage by EcoR124I has also been reported on. This should provide good proof of concept for the promise of aptamer targeting and inhibition of both

PFEIF4A and PFDH60, and should not be disregarded when considering future investigation into novel anti-malarial therapeutics and diagnostic platforms.

Another area of interest in determining the suitability of PFEIF4A as an anti-malarial target was its direct comparison with human EIF4A, and a clone was kindly provided by Dr Christian Wasmer of Prof. Gerhard Wagners' laboratory (Harvard Medical School, Boston, Massachusetts, USA). Furthermore, human EIF4A activity has been shown to be greatly influenced by other EIF4 proteins (Imataka *et al.* 1997; Rogers *et al.* 2001; Oberer *et al.* 2005; Rozovsky *et al.* 2008; Marintchev *et al.* 2009; Özeş *et al.* 2011), and more recently potential *P. falciparum* EIF4 orthologues have been identified (Tuteja 2009; Tuteja *et al.* 2009; Tuteja *et al.* 2010). Therefore, *PFEIF4E*, *PFEIF4G*, *P. falciparum* poly-A binding protein (*PFPABP*), and *P. falciparum* single-stranded binding protein (*PFSSB*) genes were synthesised by GENEART (Regensburg, Bavaria, Germany) and cloned into the pET-28a(+) expression vector. Despite preliminary expression analysis and purification of PFEIF4E and PFEIF4G, a thorough comparison could not be performed due to time restraints. However, particular interest was placed in comparing the inhibition of the *P. falciparum* and human proteins with established compounds as well as inhibitor molecules already shown to affect human EIF4A, such as hippuristanol (Lindqvist *et al.* 2008), or cytotoxins such as BPSL1549 (Cruz-Migoni *et al.* 2011). A particularly interesting outcome would be the identification of compounds that could distinguish between PFEIF4A and human EIF4A.

7.4 Summary

The increase of drug-resistant parasites and insecticide-resistant *A. gambiae* mosquitoes has resulted in the call for new approaches to combat malaria. One proposed strategy focuses on the targeting of *Plasmodium* helicases due to their necessity in all cellular processes, and therefore the organisms' survival and infection of their host.

PFEIF4A has emerged as a target of particular interest due to its similarity to human EIF4A. Although further investigation is required to fully appreciate its potential as an anti-malarial target, this study has furthered the understanding of its biochemistry, with particular interest in its absolute requirements for specific substrate topologies, and the

influence temperature has on its unwinding activity. This study has also expanded on the susceptibility of PFEIF4A to inhibition from DNA binding compounds, as well as providing preliminary information on its structure.

Significant *in vivo* studies as well as biochemical and biophysical experimentations have been performed to understand the underlying activities of PFEIF4A as an integral protein to the life processes of *P. falciparum*. However, studies into PFEIF4A as part of the *P. falciparum* EIF4 complex and novel inhibitors that distinguish it from human proteins are now required. Furthermore, structural studies into how the complex forms and where interaction by small molecules takes place would be of interest for the pursuit of developing more effective anti-malarial solutions in the future.

REFERENCES

Abbate, E. A., J. M. Berger, et al. (2004). "The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2." Genes Dev **18**(16): 1981-1996.

Adams, S., H. Brown, et al. (2002). "Breaking down the blood-brain barrier: signaling a path to cerebral malaria?" Trends Parasitol **18**(8): 360-366.

Agnandji, S. T., B. Lell, et al. (2011). "First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children." N Engl J Med **365**(20): 1863-1875.

Ahmad, A. I. (2007). "BOXTO as a real-time thermal cycling reporter dye." J Biosci **32**(2): 229-239.

Ahn, S. J., J. Costa, et al. (1996). "PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR." Nucleic Acids Res **24**(13): 2623-2625.

Aley, S. B., J. A. Sherwood, et al. (1984). "Knob-positive and knob-negative Plasmodium falciparum differ in expression of a strain-specific malarial antigen on the surface of infected erythrocytes." J Exp Med **160**(5): 1585-1590.

Ali, J. A., N. K. Maluf, et al. (1999). "An oligomeric form of E. coli UvrD is required for optimal helicase activity." J Mol Biol **293**(4): 815-834.

Alphey, L., C. B. Beard, et al. (2002). "Malaria control with genetically manipulated insect vectors." Science **298**(5591): 119-121.

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-410.

Anand, S. P. and S. A. Khan (2004). "Structure-specific DNA binding and bipolar helicase activities of PcrA." Nucleic Acids Res **32**(10): 3190-3197.

Aunpad, R., S. Somsri, et al. (2009). "The effect of mimicking febrile temperature and drug stress on malarial development." Ann Clin Microbiol Antimicrob **8**: 19.

Baker, D. A. (2010). "Malaria gametocytogenesis." Mol Biochem Parasitol **172**(2): 57-65.

Barcena, M., T. Ruiz, et al. (2001). "The DnaB.DnaC complex: a structure based on dimers assembled around an occluded channel." EMBO J **20**(6): 1462-1468.

Bennett, R. J. and J. L. Keck (2004). "Structure and function of RecQ DNA helicases." Crit Rev Biochem Mol Biol **39**(2): 79-97.

Birkholtz, L. M., G. Blatch, et al. (2008). "Heterologous expression of plasmodial proteins for structural studies and functional annotation." Malar J **7**: 197.

Birnboim, H. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic Acids Res **7**(6): 1513-1523.

Bledsoe, G. H. (2005). "Malaria primer for clinicians in the United States." South Med J **98**(12): 1197-1204; quiz 1205, 1230.

Bloland, P. B. and H. A. Williams (2002). Malaria Control During Mass Population Movements and Natural Disasters. Washington, National Academies Press.

Blum, A., J. Bottcher, et al. (2011). "Two solutions for the same problem: multiple binding modes of pyrrolidine-based HIV-1 protease inhibitors." J Mol Biol **410**(4): 745-755.

Bodenhausen, G. and D. J. Ruben (1980). "Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy." Chemical Physics Letters **69**(1): 185-189.

Brendza, K. M., W. Cheng, et al. (2005). "Autoinhibition of Escherichia coli Rep monomer helicase activity by its 2B subdomain." Proc Natl Acad Sci U S A **102**(29): 10076-10081.

Brey, P. T. (2003). "Anopheles gambiae genome: perspectives for malaria control." Mol Cells **15**(2): 133-138.

Brown, P. H. and P. Schuck (2006). "Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation." Biophys J **90**(12): 4651-4661.

Bujalowski, W. and M. M. Klonowska (1993). "Negative cooperativity in the binding of nucleotides to Escherichia coli replicative helicase DnaB protein. Interactions with fluorescent nucleotide analogs." Biochemistry **32**(22): 5888-5900.

Burnette, W. N. (1981). ""Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A." Anal Biochem **112**(2): 195-203.

Byrd, A. K. and K. D. Raney (2004). "Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA." Nat Struct Mol Biol **11**(6): 531-538.

Byrd, A. K. and K. D. Raney (2005). "Increasing the length of the single-stranded overhang enhances unwinding of duplex DNA by bacteriophage T4 Dda helicase." Biochemistry **44**(39): 12990-12997.

Caruthers, J. M., E. R. Johnson, et al. (2000). "Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase." Proc Natl Acad Sci U S A **97**(24): 13080-13085.

Caruthers, J. M. and D. B. McKay (2002). "Helicase structure and mechanism." Curr Opin Struct Biol **12**(1): 123-133.

Chan, K. M., D. Delfert, et al. (1986). "A direct colorimetric assay for Ca2+ -stimulated ATPase activity." Anal Biochem **157**(2): 375-380.

Chen, C. S., C. T. Chiou, et al. (2009). "Structure-based discovery of triphenylmethane derivatives as inhibitors of hepatitis C virus helicase." J Med Chem **52**(9): 2716-2723.

Chen, H., S. P. Meisburger, et al. (2012). "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA." Proc Natl Acad Sci U S A **109**(3): 799-804.

Chen, Q., M. Schlichtherle, et al. (2000). "Molecular aspects of severe malaria." Clin Microbiol Rev **13**(3): 439-450.

Chervenka, C. H. (1970). "Long-column meniscus depletion sedimentation equilibrium technique for the analytical ultracentrifuge." Anal Biochem **34**: 24-29.

Chong, J. P., M. K. Hayashi, et al. (2000). "A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase." Proc Natl Acad Sci U S A **97**(4): 1530-1535.

Ciampi, M. S. (2006). "Rho-dependent terminators and transcription termination." Microbiology **152**(Pt 9): 2515-2528.

Cirimotich, C. M., Y. Dong, et al. (2010). "Mosquito immune defenses against Plasmodium infection." Dev Comp Immunol **34**(4): 387-395.

Cobb, J. A., L. Bjergbaek, et al. (2002). "RecQ helicases: at the heart of genetic stability." FEBS Lett **529**(1): 43-48.

Coleman, P. G. and L. Alphey (2004). "Genetic control of vector populations: an imminent prospect." Trop Med Int Health **9**(4): 433-437.

Conroy, S. C., T. E. Dever, et al. (1990). "Characterization of the 46,000-dalton subunit of eIF-4F." Arch Biochem Biophys **282**(2): 363-371.

Cordin, O., J. Banroques, et al. (2006). "The DEAD-box protein family of RNA helicases." Gene **367**: 17-37.

Crosnier, C., L. Y. Bustamante, et al. (2011). "Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum." Nature **480**(7378): 534-537.

Cruz-Migoni, A., G. M. Hautbergue, et al. (2011). "A Burkholderia pseudomallei toxin inhibits helicase activity of translation factor eIF4A." Science **334**(6057): 821-824.

D'Angelo, E., J. Crutchfield, et al. (2001). "Rapid, sensitive, microscale determination of phosphate in water and soil." J Environ Qual **30**(6): 2206-2209.

Das, K., S. E. Martinez, et al. (2012). "HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism." Nat Struct Mol Biol **19**(2): 253-259.

Davey, M. J. and M. O'Donnell (2003). "Replicative helicase loaders: ring breakers and ring makers." Curr Biol **13**(15): R594-596.

De Clercq, E. (2002). "Strategies in the design of antiviral drugs." Nat Rev Drug Discov **1**(1): 13-25.

De Clercq, E. (2004). "Antivirals and antiviral strategies." Nat Rev Microbiol **2**(9): 704-720.

de la Cruz, J., D. Kressler, et al. (1999). "Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families." Trends Biochem Sci **24**(5): 192-198.

Deaconescu, A. M., A. L. Chambers, et al. (2006). "Structural basis for bacterial transcription-coupled DNA repair." Cell **124**(3): 507-520.

Diges, C. M. and O. C. Uhlenbeck (2001). "Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA." EMBO J **20**(19): 5503-5512.

Dillingham, M. S., D. B. Wigley, et al. (2000). "Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed." Biochemistry **39**(1): 205-212.

Djimdé, A. A., O. K. Doumbo, et al. (2003). "Clearance of drug-resistant parasites as a model for protective immunity in Plasmodium falciparum malaria." Am J Trop Med Hyg **69**(5): 558-563.

Doggrell, S. A. (2005). "Pegaptanib: the first antiangiogenic agent approved for neovascular macular degeneration." Expert Opin Pharmacother **6**(8): 1421-1423.

Dondorp, A. M., F. Nosten, et al. (2009). "Artemisinin resistance in Plasmodium falciparum malaria." N Engl J Med **361**(5): 455-467.

Doucleff, M., M. Hatcher-Skeers, et al. (2011). Pocket Guide to Biomolecular NMR.

Douglas, A. D., A. R. Williams, et al. (2011). "The blood-stage malaria antigen PfRH5 is susceptible to vaccine-inducible cross-strain neutralizing antibody." Nat Commun **2**: 601.

Du, M. X., R. B. Johnson, et al. (2002). "Comparative characterization of two DEAD-box RNA helicases in superfamily II: human translation-initiation factor 4A and hepatitis C virus non-structural protein 3 (NS3) helicase." Biochem J **363**(Pt 1): 147-155.

Dumont, S., W. Cheng, et al. (2006). "RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP." Nature **439**(7072): 105-108.

Durr, H., C. Korner, et al. (2005). "X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA." Cell **121**(3): 363-373.

Dutta, A., S. Zheng, et al. (2011). "Intermolecular Interactions within the Abundant DEAD-box Protein Dhh1 Regulate Its Activity in Vivo." J Biol Chem **286**(31): 27454-27470.

EANMAT (2003). "The efficacy of antimalarial monotherapies, sulphadoxine-pyrimethamine and amodiaquine in East Africa: implications for sub-regional policy." Trop Med Int Health **8**(10): 860-867.

Ellington, A. D. and J. W. Szostak (1990). "In vitro selection of RNA molecules that bind specific ligands." Nature **346**(6287): 818-822.

Enemark, E. J., G. Chen, et al. (2000). "Crystal structure of the DNA binding domain of the replication initiation protein E1 from papillomavirus." Mol Cell **6**(1): 149-158.

Enemark, E. J. and L. Joshua-Tor (2006). "Mechanism of DNA translocation in a replicative hexameric helicase." Nature **442**(7100): 270-275.

Eriksson, M., H. J. Karlsson, et al. (2003). "Groove-binding unsymmetrical cyanine dyes for staining of DNA: dissociation rates in free solution and electrophoresis gels." Nucleic Acids Res **31**(21): 6235-6242.

Erzberger, J. P. and J. M. Berger (2006). "Evolutionary relationships and structural mechanisms of AAA+ proteins." Annu Rev Biophys Biomol Struct **35**: 93-114.

Evans, L., D. Gowers, et al. (2012). "Enhanced purification and characterization of the PfeIF4A (PfH45) helicase from Plasmodium falciparum using a codon-optimised clone." Protein Expr Purif **85**(1): 1-8.

Fairman-Williams, M. E., U. P. Guenther, et al. (2010). "SF1 and SF2 helicases: family matters." Curr Opin Struct Biol **20**(3): 313-324.

Fairman, M. E., P. A. Maroney, et al. (2004). "Protein displacement by DExH/D "RNA helicases" without duplex unwinding." Science **304**(5671): 730-734.

Falkenberg, M., P. Elias, et al. (1998). "The herpes simplex virus type 1 helicase-primase. Analysis of helicase activity." J Biol Chem **273**(48): 32154-32157.

Farmerie, W. G., D. D. Loeb, et al. (1987). "Expression and processing of the AIDS virus reverse transcriptase in Escherichia coli." Science **236**(4799): 305-308.

Farrell, D. F. (2012). "Retinal toxicity to antimalarial drugs: chloroquine and hydroxychloroquine: a neurophysiologic study." Clin Ophthalmol **6**: 377-383.

Feng, J., Y. Chen, et al. (2011). "An improved malachite green assay of phosphate: mechanism and application." Anal Biochem **409**(1): 144-149.

Firman, K., L. Evans, et al. (2012). "A Synthetic Biology Project - Developing a single-molecule device for screening drug-target interactions." FEBS Lett **586**(15): 2157-2163.

Firman, K. and M. D. Szczelkun (2000). "Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement." EMBO J **19**(9): 2094-2102.

Fischer, C. J., N. K. Maluf, et al. (2004). "Mechanism of ATP-dependent translocation of E.coli UvrD monomers along single-stranded DNA." J Mol Biol **344**(5): 1287-1309.

Flaus, A., D. M. Martin, et al. (2006). "Identification of multiple distinct Snf2 subfamilies with conserved structural motifs." Nucleic Acids Res **34**(10): 2887-2905.

Flaus, A. and T. Owen-Hughes (2004). "Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer?" Curr Opin Genet Dev **14**(2): 165-173.

Fletcher, R. J., B. E. Bishop, et al. (2003). "The structure and function of MCM from archaeal M. Thermoautotrophicum." Nat Struct Biol **10**(3): 160-167.

Fox, J. M. and I. Erill (2010). "Relative codon adaptation: a generic codon bias index for prediction of gene expression." DNA Res **17**(3): 185-196.

Franke, D. and D. I. Svergun (2009). "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." Journal of Applied Crystallography **42**(2): 342-346.

French, N., J. Nakiyingi, et al. (2001). "Increasing rates of malarial fever with deteriorating immune status in HIV-1-infected Ugandan adults." AIDS **15**(7): 899-906.

Fried, M. and D. M. Crothers (1981). "Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis." Nucleic Acids Res **9**(23): 6505-6525.

Gai, D., R. Zhao, et al. (2004). "Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen." Cell **119**(1): 47-60.

Galletto, R., M. J. Jezewska, et al. (2004). "Unzipping mechanism of the double-stranded DNA unwinding by a hexameric helicase: quantitative analysis of the rate of the dsDNA unwinding, processivity and kinetic step-size of the Escherichia coli DnaB helicase using rapid quench-flow method." J Mol Biol **343**(1): 83-99.

García-Gómez, J. J., S. Lebaron, et al. (2011). "Dynamics of the putative RNA helicase Spb4 during ribosome assembly in Saccharomyces cerevisiae." Mol Cell Biol.

Gardner, M. J., N. Hall, et al. (2002). "Genome sequence of the human malaria parasite Plasmodium falciparum." Nature **419**(6906): 498-511.

Garner, M. M. and A. Revzin (1981). "A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system." Nucleic Acids Res **9**(13): 3047-3060.

Gogol, E. P., S. E. Seifried, et al. (1991). "Structure and assembly of the Escherichia coli transcription termination factor rho and its interaction with RNA. I. Cryoelectron microscopic studies." J Mol Biol **221**(4): 1127-1138.

Goldberg, A. L. (2003). "Protein degradation and protection against misfolded or damaged proteins." Nature **426**(6968): 895-899.

Gorbalenya, A. E. and E. V. Koonin (1993). "Helicases: amino acid sequence comparisons and structure-function relationships." Current Opinion in Structural Biology **3**(3): 419-429.

Gowthaman, R., D. Sekhar, et al. (2005). "A database for Plasmodium falciparum protein models." Bioinformation **1**(2): 50-51.

Greenwood, B. and T. Mutabingwa (2002). "Malaria in 2002." Nature **415**(6872): 670-672.

Greenwood, B. M., K. Bojang, et al. (2005). "Malaria." Lancet **365**(9469): 1487-1498.

Greenwood, T., T. Vikerfors, et al. (2008). "Febrile Plasmodium falciparum malaria 4 years after exposure in a man with sickle cell disease." Clin Infect Dis **47**(4): e39-41.

Gruber, A. R., R. Lorenz, et al. (2008). "The Vienna RNA websuite." Nucleic Acids Res **36**(Web Server issue): W70-74.

Grune, T., J. Brzeski, et al. (2003). "Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI." Mol Cell **12**(2): 449-460.

Gupta, A., P. Mehra, et al. (2006). "Analogous expression pattern of Plasmodium falciparum replication initiation proteins PfMCM4 and PfORC1 during the asexual and sexual stages of intraerythrocytic developmental cycle." FEMS Microbiol Lett **261**(1): 12-18.

Ha, T., I. Rasnik, et al. (2002). "Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase." Nature **419**(6907): 638-641.

Hajduk, P. J., J. R. Huth, et al. (2005). "Druggability indices for protein targets derived from NMR-based screening data." J Med Chem **48**(7): 2518-2525.

Hall, M. C. and S. W. Matson (1999). "Helicase motifs: the engine that powers DNA unwinding." Mol Microbiol **34**(5): 867-877.

Hallem, E. A., A. Nicole Fox, et al. (2004). "Olfaction: mosquito receptor for human-sweat odorant." Nature **427**(6971): 212-213.

Hallett, R. L., C. J. Sutherland, et al. (2004). "Combination therapy counteracts the enhanced transmission of drug-resistant malaria parasites to mosquitoes." Antimicrob Agents Chemother **48**(10): 3940-3943.

Hanahan, D., J. Jessee, et al. (1991). "Plasmid transformation of Escherichia coli and other bacteria." Methods Enzymol **204**: 63-113.

Harris, C., I. Morlais, et al. (2012). "Plasmodium falciparum Produce Lower Infection Intensities in Local versus Foreign Anopheles gambiae Populations." PLoS One **7**(1): e30849.

Hastings, I. M., W. M. Watkins, et al. (2002). "The evolution of drug-resistant malaria: the role of drug elimination half-life." Philos Trans R Soc Lond B Biol Sci **357**(1420): 505-519.

Hay, S. I., D. J. Rogers, et al. (2002). "Hot topic or hot air? Climate change and malaria resurgence in East African highlands." Trends Parasitol **18**(12): 530-534.

Hickman, A. B. and F. Dyda (2005). "Binding and unwinding: SF3 viral helicases." Curr Opin Struct Biol **15**(1): 77-85.

Hickman, A. B., D. R. Ronning, et al. (2002). "Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep." Mol Cell **10**(2): 327-337.

Hingorani, M. M., M. T. Washington, et al. (1997). "The dTTPase mechanism of T7 DNA helicase resembles the binding change mechanism of the F1-ATPase." Proc Natl Acad Sci U S A **94**(10): 5012-5017.

Hizi, A., C. McGill, et al. (1988). "Expression of soluble, enzymatically active, human immunodeficiency virus reverse transcriptase in Escherichia coli and analysis of mutants." Proc Natl Acad Sci U S A **85**(4): 1218-1222.

Hochuli, E. (1990). "Purification of recombinant proteins with metal chelate adsorbent." Genet Eng (N Y) **12**: 87-98.

Holladay, L. A. (1980). "Simultaneous rapid estimation of sedimentation coefficient and molecular weight." Biophys Chem **11**(2): 303-308.

Holt, R. A., G. M. Subramanian, et al. (2002). "The genome sequence of the malaria mosquito Anopheles gambiae." Science **298**(5591): 129-149.

Ilyina, T. V., A. E. Gorbalenya, et al. (1992). "Organization and evolution of bacterial and bacteriophage primase-helicase systems." J Mol Evol **34**(4): 351-357.

Imataka, H. and N. Sonenberg (1997). "Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A." Mol Cell Biol **17**(12): 6940-6947.

Ivessa, A. S., J. Q. Zhou, et al. (2002). "Saccharomyces Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and subtelomeric DNA." Genes Dev **16**(11): 1383-1396.

Iyer, L. M., D. D. Leipe, et al. (2004). "Evolutionary history and higher order classification of AAA+ ATPases." J Struct Biol **146**(1-2): 11-31.

James, J. A., A. K. Aggarwal, et al. (2004). "Structure of adeno-associated virus type 2 Rep40-ADP complex: insight into nucleotide recognition and catalysis by superfamily 3 helicases." Proc Natl Acad Sci U S A **101**(34): 12455-12460.

James, J. A., C. R. Escalante, et al. (2003). "Crystal structure of the SF3 helicase from adeno-associated virus type 2." Structure **11**(8): 1025-1035.

Jankowsky, E., C. H. Gross, et al. (2001). "Active disruption of an RNA-protein interaction by a DExH/D RNA helicase." Science **291**(5501): 121-125.

Janscak, P. and T. A. Bickle (1998). "The DNA recognition subunit of the type IB restriction-modification enzyme EcoAI tolerates circular permutions of its polypeptide chain." J Mol Biol **284**(4): 937-948.

Jayasena, S. D. (1999). "Aptamers: an emerging class of molecules that rival antibodies in diagnostics." Clin Chem **45**(9): 1628-1650.

Jennings, E. T. (2006). Curing the Colonizers: Hydrotherapy, Climatology, and French Colonial Spas, Duke University Press.

Jensen, A. T., P. Magistrado, et al. (2004). "Plasmodium falciparum associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A var genes." J Exp Med **199**(9): 1179-1190.

Johnson, M., I. Zaretskaya, et al. (2008). "NCBI BLAST: a better web interface." Nucleic Acids Res **36**(Web Server issue): W5-9.

Kaczor, A. and D. Matosiuk (2010). "Structure-based virtual screening for novel inhibitors of Japanese encephalitis virus NS3 helicase/nucleoside triphosphatase." FEMS Immunol Med Microbiol **58**(1): 91-101.

Kamatari, Y. O., T. Konno, et al. (1998). "The methanol-induced transition and the expanded helical conformation in hen lysozyme." Protein Sci **7**(3): 681-688.

Kanai, Y., N. Dohmae, et al. (2004). "Kinesin transports RNA: isolation and characterization of an RNA-transporting granule." Neuron **43**(4): 513-525.

Kaplan, D. L. and M. O'Donnell (2004). "Twin DNA pumps of a hexameric helicase provide power to simultaneously melt two duplexes." Mol Cell **15**(3): 453-465.

Kappe, S. H., A. M. Vaughan, et al. (2010). "That was then but this is now: malaria research in the time of an eradication agenda." Science **328**(5980): 862-866.

Karlsson, H. J., M. Eriksson, et al. (2003). "Groove-binding unsymmetrical cyanine dyes for staining of DNA: syntheses and characterization of the DNA-binding." Nucleic Acids Res **31**(21): 6227-6234.

Keeler, J. (2011). Understanding NMR Spectroscopy, John Wiley & Sons.

Kelley, L. A. and M. J. Sternberg (2009). "Protein structure prediction on the Web: a case study using the Phyre server." Nat Protoc **4**(3): 363-371.

Kiianitsa, K., J. A. Solinger, et al. (2003). "NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities." Anal Biochem **321**(2): 266-271.

Kim, D. E., M. Narayan, et al. (2002). "T7 DNA helicase: a molecular motor that processively and unidirectionally translocates along single-stranded DNA." J Mol Biol **321**(5): 807-819.

Kim, J. L., K. A. Morgenstern, et al. (1998). "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding." Structure **6**(1): 89-100.

Kim, J. W., M. Y. Seo, et al. (2003). "Structurally conserved amino Acid w501 is required for RNA helicase activity but is not essential for DNA helicase activity of hepatitis C virus NS3 protein." J Virol **77**(1): 571-582.

Komori, K., R. Fujikane, et al. (2002). "Novel endonuclease in Archaea cleaving DNA with various branched structure." Genes Genet Syst **77**(4): 227-241.

Konarev, P. V., M. V. Petoukhov, et al. (2006). "ATSAS 2.1, a program package for small-angle scattering data analysis." Journal of Applied Crystallography **39**(2): 277-286.

Konarev, P. V., V. V. Volkov, et al. (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." Journal of Applied Crystallography **36**(5): 1277-1282.

Koonin, E. V. (1993). "A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication." Nucleic Acids Res **21**(11): 2541-2547.

Korangy, F. and D. A. Julin (1993). "Kinetics and processivity of ATP hydrolysis and DNA unwinding by the RecBC enzyme from Escherichia coli." Biochemistry **32**(18): 4873-4880.

Korenromp, E. L., B. G. Williams, et al. (2003). "Measurement of trends in childhood malaria mortality in Africa: an assessment of progress toward targets based on verbal autopsy." Lancet Infect Dis **3**(6): 349-358.

Korhonen, J. A., M. Gaspari, et al. (2003). "TWINKLE Has 5' -> 3' DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein." J Biol Chem **278**(49): 48627-48632.

Korolev, S., J. Hsieh, et al. (1997). "Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP." Cell **90**(4): 635-647.

Korolev, S., N. Yao, et al. (1998). "Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases." Protein Sci **7**(3): 605-610.

Kotecha, M., J. Kluza, et al. (2008). "Inhibition of DNA binding of the NF-Y transcription factor by the pyrrolobenzodiazepine-polyamide conjugate GWL-78." Mol Cancer Ther **7**(5): 1319-1328.

Kovelman, R. and R. G. Roeder (1990). "Sarkosyl defines three intermediate steps in transcription initiation by RNA polymerase III: application to stimulation of transcription by E1A." Genes Dev **4**(4): 646-658.

Kowalczykowski, S. C. (2000). "Initiation of genetic recombination and recombination-dependent replication." Trends Biochem Sci **25**(4): 156-165.

Kozin, M. B. and D. I. Svergun (2001). "Automated matching of high- and low-resolution structural models." Journal of Applied Crystallography **34**(1): 33-41.

Kremsner, P. G. and S. Krishna (2004). "Antimalarial combinations." Lancet **364**(9430): 285-294.

Kublin, J. G., P. Patnaik, et al. (2005). "Effect of Plasmodium falciparum malaria on concentration of HIV-1-RNA in the blood of adults in rural Malawi: a prospective cohort study." Lancet **365**(9455): 233-240.

Kuehn, A., N. Simon, et al. (2010). "Family members stick together: multi-protein complexes of malaria parasites." Med Microbiol Immunol **199**(3): 209-226.

Kwiatkowski, D. (1989). "Febrile temperatures can synchronize the growth of Plasmodium falciparum in vitro." J Exp Med **169**(1): 357-361.

Kwong, A. D., B. G. Rao, et al. (2005). "Viral and cellular RNA helicases as antiviral targets." Nat Rev Drug Discov **4**(10): 845-853.

Labib, K., J. A. Tercero, et al. (2000). "Uninterrupted MCM2-7 function required for DNA replication fork progression." Science **288**(5471): 1643-1647.

Lam, A. M., R. S. Rypma, et al. (2004). "Enhanced nucleic acid binding to ATP-bound hepatitis C virus NS3 helicase at low pH activates RNA unwinding." Nucleic Acids Res **32**(13): 4060-4070.

Larder, B., D. Purifoy, et al. (1987). "AIDS virus reverse transcriptase defined by high level expression in Escherichia coli." EMBO J **6**(10): 3133-3137.

Larkin, M. A., G. Blackshields, et al. (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Latt, S. A., G. Stetten, et al. (1975). "Recent developments in the detection of deoxyribonucleic acid synthesis by 33258 Hoechst fluorescence." J Histochem Cytochem **23**(7): 493-505.

Levin, M. K., Y. H. Wang, et al. (2004). "The functional interaction of the hepatitis C virus helicase molecules is responsible for unwinding processivity." J Biol Chem **279**(25): 26005-26012.

Li, D., R. Zhao, et al. (2003). "Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen." Nature **423**(6939): 512-518.

Linder, P. and P. Lasko (2006). "Bent out of shape: RNA unwinding by the DEAD-box helicase Vasa." Cell **125**(2): 219-221.

Linder, P., P. F. Lasko, et al. (1989). "Birth of the D-E-A-D box." Nature **337**(6203): 121-122.

Linder, P., N. K. Tanner, et al. (2001). "From RNA helicases to RNPases." Trends Biochem Sci **26**(6): 339-341.

Lindqvist, L., M. Oberer, et al. (2008). "Selective pharmacological targeting of a DEAD box RNA helicase." PLoS One **3**(2): e1583.

Litsios, S. (1996). The Tomorrow of Malaria. Wellington, Pacific Press.

Lohman, T. M. and K. P. Bjornson (1996). "Mechanisms of helicase-catalyzed DNA unwinding." Annu Rev Biochem **65**: 169-214.

Long, D. T. and K. N. Kreuzer (2009). "Fork regression is an active helicase-driven pathway in bacteriophage T4." EMBO Rep **10**(4): 394-399.

Long, H. Y., B. Lell, et al. (2001). "Plasmodium falciparum: in vitro growth inhibition by febrile temperatures." Parasitol Res **87**(7): 553-555.

Loomis, W. F., Jr. and B. Magasanik (1967). "Glucose-lactose diauxie in Escherichia coli." J Bacteriol **93**(4): 1397-1401.

Luo, X., D. G. Sanford, et al. (1996). "Solution structure of the origin DNA-binding domain of SV40 T-antigen." Nat Struct Biol **3**(12): 1034-1039.

Mackintosh, S. G., J. Z. Lu, et al. (2006). "Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the hepatitis C virus." J Biol Chem **281**(6): 3528-3535.

Mahdi, A. A., G. S. Briggs, et al. (2003). "A model for dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins." EMBO J **22**(3): 724-734.

Maier, A. G., B. M. Cooke, et al. (2009). "Malaria parasite proteins that remodel the host erythrocyte." Nat Rev Microbiol **7**(5): 341-354.

Manosas, M., X. G. Xi, et al. (2010). "Active and passive mechanisms of helicases." Nucleic Acids Res **38**(16): 5518-5526.

Marintchev, A., K. A. Edmonds, et al. (2009). "Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation." Cell **136**(3): 447-460.

Marsden, S., M. Nardelli, et al. (2006). "Unwinding single RNA molecules using helicases involved in eukaryotic translation initiation." J Mol Biol **361**(2): 327-335.

Martin, A., T. A. Baker, et al. (2005). "Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines." Nature **437**(7062): 1115-1120.

Matson, S. W. (1991). "DNA helicases of Escherichia coli." Prog Nucleic Acid Res Mol Biol **40**: 289-326.

Matson, S. W., D. W. Bean, et al. (1994). "DNA helicases: enzymes with essential roles in all aspects of DNA metabolism." Bioessays **16**(1): 13-22.

Matson, S. W. and K. A. Kaiser-Rogers (1990). "DNA helicases." Annu Rev Biochem **59**: 289-329.

Mayxay, M., M. Khanthavong, et al. (2004). "Randomized comparison of chloroquine plus sulfadoxine-pyrimethamine versus artesunate plus mefloquine versus artemether-lumefantrine in the treatment of uncomplicated falciparum malaria in the Lao People's Democratic Republic." Clin Infect Dis **39**(8): 1139-1147.

McClelland, S. E., D. T. Dryden, et al. (2005). "Continuous assays for DNA translocation using fluorescent triplex dissociation: application to type I restriction endonucleases." J Mol Biol **348**(4): 895-915.

Mehlhorn, H. and W. Peters (1980). "The formation of kinetes and oocysts in plasmodium gallinaceum and considerations on phylogenetic relationships between haemosporidia, piroplasmida, and other coccidia." Protistologica **16**(1): 135-154.

Mehta, J. and R. Tuteja (2011). "A novel dual Dbp5/DDX19 homologue from Plasmodium falciparum requires Q motif for activity." Mol Biochem Parasitol **176**(1): 58-63.

Miller, L. H., M. F. Good, et al. (1994). "Malaria pathogenesis." Science **264**(5167): 1878-1883.

Mita, T., A. Kaneko, et al. (2003). "Recovery of chloroquine sensitivity and low prevalence of the Plasmodium falciparum chloroquine resistance transporter gene mutation K76T following the discontinuance of chloroquine use in Malawi." Am J Trop Med Hyg **68**(4): 413-415.

Morgan, R. E., G. O. Batot, et al. (2011). "Crystal structures of Burkholderia cenocepacia dihydropteroate synthase in the apo-form and complexed with the product 7,8-dihydropteroate." BMC Struct Biol **11**: 21.

Mount, A. M., V. Mwapasa, et al. (2004). "Impairment of humoral immunity to Plasmodium falciparum malaria in pregnancy by HIV infection." Lancet **363**(9424): 1860-1867.

Mous, J., E. P. Heimer, et al. (1988). "Processing protease and reverse transcriptase from human immunodeficiency virus type I polyprotein in Escherichia coli." J Virol **62**(4): 1433-1436.

Müller, B., T. Restle, et al. (1989). "Co-expression of the subunits of the heterodimer of HIV-1 reverse transcriptase in Escherichia coli." J Biol Chem **264**(24): 13975-13978.

Mutabingwa, T. K., D. Anthony, et al. (2005). "Amodiaquine alone, amodiaquine+sulfadoxine-pyrimethamine, amodiaquine+artesunate, and artemether-lumefantrine for outpatient treatment of malaria in Tanzanian children: a four-arm randomised effectiveness trial." Lancet **365**(9469): 1474-1480.

Myers, E. W. and W. Miller (1988). "Optimal alignments in linear space." Comput Appl Biosci **4**(1): 11-17.

Nanduri, B., A. K. Byrd, et al. (2002). "Pre-steady-state DNA unwinding by bacteriophage T4 Dda helicase reveals a monomeric molecular motor." Proc Natl Acad Sci U S A **99**(23): 14722-14727.

Newton, C. R., T. T. Hien, et al. (2000). "Cerebral malaria." J Neurol Neurosurg Psychiatry **69**(4): 433-441.

Niedenzu, T., D. Roleke, et al. (2001). "Crystal structure of the hexameric replicative helicase RepA of plasmid RSF1010." J Mol Biol **306**(3): 479-487.

Oberer, M., A. Marintchev, et al. (2005). "Structural basis for the enhancement of eIF4A helicase activity by eIF4G." Genes Dev **19**(18): 2212-2223.

Ortega, A., D. Amoros, et al. (2011). "Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models." Biophys J **101**(4): 892-898.

Özeş, A. R., K. Feoktistova, et al. (2011). "Duplex unwinding and ATPase activities of the DEAD-box helicase eIF4A are coupled by eIF4G and eIF4B." J Mol Biol **412**(4): 674-687.

Özsoy, A. Z., H. M. Ragonese, et al. (2003). "Analysis of helicase activity and substrate specificity of Drosophila RECQ5." Nucleic Acids Res **31**(5): 1554-1564.

Palchaudhuri, R. and P. J. Hergenrother (2007). "DNA as a target for anticancer compounds: methods to determine the mode of binding and the mechanism of action." Curr Opin Biotechnol **18**(6): 497-503.

Pang, P. S., E. Jankowsky, et al. (2002). "The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding." EMBO J **21**(5): 1168-1176.

Paolini, C., A. Lahm, et al. (2000). "Mutational analysis of hepatitis C virus NS3-associated helicase." J Gen Virol **81**(Pt 7): 1649-1658.

Pape, T., H. Meka, et al. (2003). "Hexameric ring structure of the full-length archaeal MCM protein complex." EMBO Rep **4**(11): 1079-1083.

Pasternak, N. D. and R. Dzikowski (2009). "PfEMP1: an antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite Plasmodium falciparum." Int J Biochem Cell Biol **41**(7): 1463-1466.

Patel, J., I. Taylor, et al. (1992). "High-level expression of the cloned genes encoding the subunits of and intact DNA methyltransferase, M.EcoR124." Gene **112**(1): 21-27.

Patterson, S., M. S. Alphey, et al. (2011). "Dihydroquinazolines as a Novel Class of Trypanosoma brucei Trypanothione Reductase Inhibitors: Discovery, Synthesis, and Characterization of their Binding Mode by Protein Crystallography." J Med Chem.

Pause, A. and N. Sonenberg (1992). "Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A." EMBO J **11**(7): 2643-2654.

Piola, P., C. Fogg, et al. (2005). "Supervised versus unsupervised intake of six-dose artemether-lumefantrine for treatment of acute, uncomplicated Plasmodium falciparum malaria in Mbarara, Uganda: a randomised trial." Lancet **365**(9469): 1467-1473.

Plowe, C. V., J. G. Kublin, et al. (2004). "Sustained clinical efficacy of sulfadoxine-pyrimethamine for uncomplicated falciparum malaria in Malawi after 10 years as first line treatment: five year prospective study." BMJ **328**(7439): 545.

Porath, J., J. Carlsson, et al. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation." Nature **258**(5536): 598-599.

Pradhan, A., V. S. Chauhan, et al. (2005). "A novel 'DEAD-box' DNA helicase from Plasmodium falciparum is homologous to p68." Mol Biochem Parasitol **140**(1): 55-60.

Pradhan, A., V. S. Chauhan, et al. (2005). "Plasmodium falciparum DNA helicase 60 is a schizont stage specific, bipolar and dual helicase stimulated by PKC phosphorylation." Mol Biochem Parasitol **144**(2): 133-141.

Pradhan, A., V. S. Chauhan, et al. (2007). "'DEAD-box' helicase from Plasmodium falciparum is active at wide pH and is schizont stage-specific." J Vector Borne Dis **44**(1): 12-22.

Pradhan, A., E. M. Hussain, et al. (2008). "Characterization of replication fork and phosphorylation stimulated Plasmodium falciparum helicase 45." Gene **420**(1): 66-75.

Pradhan, A. and R. Tuteja (2006). "Plasmodium falciparum DNA helicase 60. dsRNA- and antibody-mediated inhibition of malaria parasite growth and downregulation of its enzyme activities by DNA-interacting compounds." FEBS J **273**(15): 3545-3556.

Pradhan, A. and R. Tuteja (2007). "Bipolar, Dual Plasmodium falciparum helicase 45 expressed in the intraerythrocytic developmental cycle is required for parasite growth." J Mol Biol **373**(2): 268-281.

Prakash, K. and R. Tuteja (2010). "A novel DEAD box helicase Has1p from Plasmodium falciparum: N-terminal is essential for activity." Parasitol Int.

Putnam, C. D., S. B. Clancy, et al. (2001). "Structure and mechanism of the RuvB Holliday junction branch migration motor." J Mol Biol **311**(2): 297-310.

Putnam, C. D., M. Hammel, et al. (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." Q Rev Biophys **40**(3): 191-285.

Py, B., C. F. Higgins, et al. (1996). "A DEAD-box RNA helicase in the Escherichia coli RNA degradosome." Nature **381**(6578): 169-172.

Rahman, K. M., C. H. James, et al. (2011). "Observation of the reversibility of a covalent pyrrolobenzodiazepine (PBD) DNA adduct by HPLC/MS and CD spectroscopy." Org Biomol Chem **9**(5): 1632-1641.

Richman, D. D. (1996). "The implications of drug resistance for strategies of combination antiviral chemotherapy." Antiviral Res **29**(1): 31-33.

Rodrigues, J., F. A. Brayner, et al. (2010). "Hemocyte differentiation mediates innate immune memory in Anopheles gambiae mosquitoes." Science **329**(5997): 1353-1355.

Rogers, G. W., Jr., A. A. Komar, et al. (2002). "eIF4A: the godfather of the DEAD box helicases." Prog Nucleic Acid Res Mol Biol **72**: 307-331.

Rogers, G. W., Jr., W. F. Lima, et al. (2001). "Further characterization of the helicase activity of eIF4A. Substrate specificity." J Biol Chem **276**(16): 12598-12608.

Rogers, G. W., Jr., N. J. Richter, et al. (2001). "Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F." J Biol Chem **276**(33): 30914-30922.

Rogers, G. W., Jr., N. J. Richter, et al. (1999). "Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A." J Biol Chem **274**(18): 12236-12244.

Rozovsky, N., A. C. Butterworth, et al. (2008). "Interactions between eIF4AI and its accessory factors eIF4B and eIF4H." RNA **14**(10): 2136-2148.

Salcedo-Amaya, A. M., W. A. Hoeijmakers, et al. (2010). "Malaria: could its unusual epigenome be the weak spot?" Int J Biochem Cell Biol **42**(6): 781-784.

Sambrook, J. and D. W. Russell (2001). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press.

Sawaya, M. R., S. Guo, et al. (1999). "Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7." Cell **99**(2): 167-177.

Scheffzek, K., M. R. Ahmadian, et al. (1997). "The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants." Science **277**(5324): 333-338.

Schneider, C. A., W. S. Rasband, et al. (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Meth **9**(7): 671-675.

Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling." Biophys J **78**(3): 1606-1619.

Schuck, P. (2005). Diffusion-deconvoluted sedimentation coefficient distributions for the analysis of interacting and non-interacting protein mixtures. Modern Analytical Ultracentrifugation: Techniques and Methods. D. J. Scott, S. E. Harding and A. J. Rowe. Cambridge, The Royal Society of Chemistry**:** 26-60.

Schuck, P. (2010). "Diffusion of the reaction boundary of rapidly interacting macromolecules in sedimentation velocity." Biophys J **98**(11): 2741-2751.

Sengoku, T., O. Nureki, et al. (2006). "Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa." Cell **125**(2): 287-300.

Seow, F., S. Sato, et al. (2005). "The plastidic DNA replication enzyme complex of Plasmodium falciparum." Mol Biochem Parasitol **141**(2): 145-153.

Serebrov, V. and A. M. Pyle (2004). "Periodic cycles of RNA unwinding and pausing by hepatitis C virus NS3 helicase." Nature **430**(6998): 476-480.

Shankar, J. and R. Tuteja (2008). "UvrD helicase of Plasmodium falciparum." Gene **410**(2): 223-233.

Singh, B., L. Kim Sung, et al. (2004). "A large focus of naturally acquired Plasmodium knowlesi infections in human beings." Lancet **363**(9414): 1017-1024.

Singleton, M. R., M. S. Dillingham, et al. (2004). "Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks." Nature **432**(7014): 187-193.

Singleton, M. R., M. S. Dillingham, et al. (2007). "Structure and mechanism of helicases and nucleic acid translocases." Annu Rev Biochem **76**: 23-50.

Singleton, M. R., M. R. Sawaya, et al. (2000). "Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides." Cell **101**(6): 589-600.

Singleton, M. R., S. Scaife, et al. (2001). "Structural analysis of DNA replication fork reversal by RecG." Cell **107**(1): 79-89.

Skordalakes, E. and J. M. Berger (2003). "Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading." Cell **114**(1): 135-146.

Snow, R. W., J. F. Trape, et al. (2001). "The past, present and future of childhood malaria mortality in Africa." Trends Parasitol **17**(12): 593-597.

Soultanas, P., M. S. Dillingham, et al. (1999). "DNA binding mediates conformational changes and metal ion coordination in the active site of PcrA helicase." J Mol Biol **290**(1): 137-148.

Soultanas, P., M. S. Dillingham, et al. (2000). "Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism." EMBO J **19**(14): 3799-3810.

Soultanas, P. and D. B. Wigley (2000). "DNA helicases: 'inching forward'." Curr Opin Struct Biol **10**(1): 124-128.

Soultanas, P. and D. B. Wigley (2001). "Unwinding the 'Gordian knot' of helicase action." Trends Biochem Sci **26**(1): 47-54.

Spencer, J., J. Amin, et al. (2011). "Synthesis and evaluation of metallocene containing methylidene-1,3-dihydro-2H-indol-2-ones as kinase inhibitors." Metallomics **3**(6): 600-608.

Staalsoe, T., C. E. Shulman, et al. (2004). "Variant surface antigen-specific IgG and protection against clinical consequences of pregnancy-associated Plasmodium falciparum malaria." Lancet **363**(9405): 283-289.

Staley, J. P. and C. Guthrie (1998). "Mechanical devices of the spliceosome: motors, clocks, springs, and things." Cell **92**(3): 315-326.

Stanley, L. K., R. Seidel, et al. (2006). "When a helicase is not a helicase: dsDNA tracking by the motor protein EcoR124I." EMBO J **25**(10): 2230-2239.

Stano, N. M., Y. J. Jeong, et al. (2005). "DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase." Nature **435**(7040): 370-373.

Stoltenburg, R., C. Reinemann, et al. (2007). "SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands." Biomol Eng **24**(4): 381-403.

Stothard, P. (2000). "The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences." Biotechniques **28**(6): 1102, 1104.

Sturm, A., R. Amino, et al. (2006). "Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids." Science **313**(5791): 1287-1290.

Subramanya, H. S., L. E. Bird, et al. (1996). "Crystal structure of a DExx box DNA helicase." Nature **384**(6607): 379-383.

Sulkowski, E. (1989). "The saga of IMAC and MIT." Bioessays **10**(5): 170-175.

Suntornthiticharoen, P., S. Petmitr, et al. (2006). "Purification and characterization of a novel 3'-5' DNA helicase from Plasmodium falciparum and its sensitivity to anthracycline antibiotics." Parasitology **133**(Pt 4): 389-398.

Sutherland, C. J., A. Alloueche, et al. (2002). "Gambian children successfully treated with chloroquine can harbor and transmit Plasmodium falciparum gametocytes carrying resistance genes." Am J Trop Med Hyg **67**(6): 578-585.

Svergun, D. (1992). "Determination of the regularization parameter in indirect-transform methods using perceptual criteria." Journal of Applied Crystallography **25**(4): 495-503.

Svergun, D. I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." Biophys J **76**(6): 2879-2886.

Szmitko, P. E., M. L. Kohn, et al. (2009). "Plasmodium falciparum malaria occurring 8 years after leaving an endemic area." Diagn Microbiol Infect Dis **63**(1): 105-107.

Tackett, A. J., Y. Chen, et al. (2005). "Multiple full-length NS3 molecules are required for optimal unwinding of oligonucleotide DNA in vitro." J Biol Chem **280**(11): 10797-10806.

Tahmaseb, K. and S. W. Matson (2010). "Rapid purification of helicase proteins and in vitro analysis of helicase activity." Methods **51**(3): 322-328.

Tai, C. L., W. K. Chi, et al. (1996). "The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3)." J Virol **70**(12): 8477-8484.

Tai, C. L., W. C. Pan, et al. (2001). "Structure-based mutational analysis of the hepatitis C virus NS3 helicase." J Virol **75**(17): 8289-8297.

Tanese, N., J. Sodroski, et al. (1986). "Expression of reverse transcriptase activity of human T-lymphotropic virus type III (HTLV-III/LAV) in Escherichia coli." J Virol **59**(3): 743-745.

Tanner, N. K., O. Cordin, et al. (2003). "The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis." Mol Cell **11**(1): 127-138.

Tanser, F. C., B. Sharp, et al. (2003). "Potential effect of climate change on malaria transmission in Africa." Lancet **362**(9398): 1792-1798.

Tarun, A. S., K. Baer, et al. (2006). "Quantitative isolation and in vivo imaging of malaria parasite liver stages." Int J Parasitol **36**(12): 1283-1293.

Taylor, A. F. and G. R. Smith (1995). "Monomeric RecBCD enzyme binds and unwinds DNA." J Biol Chem **270**(41): 24451-24458.

Taylor, I., J. Patel, et al. (1992). "Purification and biochemical characterisation of the EcoR124 type I modification methylase." Nucleic Acids Res **20**(2): 179-186.

Theunissen, C., P. Janssens, et al. (2009). "Falciparum malaria in patient 9 years after leaving malaria-endemic area." Emerg Infect Dis **15**(1): 115-116.

Thoma, N. H., B. K. Czyzewski, et al. (2005). "Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54." Nat Struct Mol Biol **12**(4): 350-356.

Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-4680.

Toth, E. A., Y. Li, et al. (2003). "The crystal structure of the bifunctional primase-helicase of bacteriophage T7." Mol Cell **12**(5): 1113-1123.

Towbin, H., T. Staehelin, et al. (1979). "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications." Proc Natl Acad Sci U S A **76**(9): 4350-4354.

Tuerk, C. and L. Gold (1990). "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science **249**(4968): 505-510.

Tuteja, R. (2007). "Helicases - feasible antimalarial drug target for Plasmodium falciparum." FEBS J **274**(18): 4699-4704.

Tuteja, R. (2007). "Malaria - an overview." FEBS J **274**(18): 4670-4679.

Tuteja, R. (2009). "Identification and bioinformatics characterization of translation initiation complex eIF4F components and poly(A)-binding protein from Plasmodium falciparum." Commun Integr Biol **2**(3): 245-260.

Tuteja, R. (2010). "Genome wide identification of Plasmodium falciparum helicases: a comparison with human host." Cell Cycle **9**(1): 104-120.

Tuteja, R. (2010). A method to inhibit the growth of Plasmodium falciparum by double-stranded RNA-mediated gene silencing of helicases. Methods Mol Biol. **587:** 389-399.

Tuteja, R. and A. Pradhan (2006). "Unraveling the 'DEAD-box' helicases of Plasmodium falciparum." Gene **376**(1): 1-12.

Tuteja, R. and A. Pradhan (2009). "Isolation and functional characterization of eIF4F components and poly(A)-binding protein from Plasmodium falciparum." Parasitol Int **58**(4): 481-485.

Tuteja, R. and A. Pradhan (2010). "PfeIF4E and PfeIF4A colocalize and their double-stranded RNA inhibits Plasmodium falciparum proliferation." Commun Integr Biol **3**(6): 611-613.

Velankar, S. S., P. Soultanas, et al. (1999). "Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism." Cell **97**(1): 75-84.

Venema, J. and D. Tollervey (1995). "Processing of pre-ribosomal RNA in Saccharomyces cerevisiae." Yeast **11**(16): 1629-1650.

Vernick, K. D. and A. P. Waters (2004). "Genomics and malaria control." N Engl J Med **351**(18): 1901-1904.

Walker, J. E., M. Saraste, et al. (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." EMBO J **1**(8): 945-951.

Washington, M. T., A. H. Rosenberg, et al. (1996). "Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding." J Biol Chem **271**(43): 26825-26834.

Webb, M. R., J. L. Plank, et al. (2007). "The phage T4 protein UvsW drives Holliday junction branch migration." J Biol Chem **282**(47): 34401-34411.

Wells, G., C. R. Martin, et al. (2006). "Design, synthesis, and biophysical and biological evaluation of a series of pyrrolobenzodiazepine-poly(N-methylpyrrole) conjugates." J Med Chem **49**(18): 5442-5461.

West, S. C. (1996). "The RuvABC proteins and Holliday junction processing in Escherichia coli." J Bacteriol **178**(5): 1237-1241.

Whitworth, J., D. Morgan, et al. (2000). "Effect of HIV-1 and increasing immunosuppression on malaria parasitaemia and clinical episodes in adults in rural Uganda: a cohort study." Lancet **356**(9235): 1051-1056.

WHO (2008). World Malaria Report 2008. Geneva, World Health Organization.

WHO (2009). World Malaria Report 2009. Geneva, World Health Organization.

WHO (2010). WHO Policy recommendation on Intermittent Preventative Treatment during infancy with sulphadoxine-pyrimethamine (SP-IPTi) for *Plasmodium falciparum* malaria control in Africa. Geneva, World Health Organization.

Wilkins, M. R., E. Gasteiger, et al. (1999). "Protein identification and analysis tools in the ExPASy server." Methods Mol Biol **112**: 531-552.

Withers, I. M., M. P. Mazanetz, et al. (2008). "Active site pressurization: a new tool for structure-guided drug design and other studies of protein flexibility." J Chem Inf Model **48**(7): 1448-1454.

Yamada, K., N. Kunishima, et al. (2001). "Crystal structure of the Holliday junction migration motor protein RuvB from Thermus thermophilus HB8." Proc Natl Acad Sci U S A **98**(4): 1442-1447.

Yao, N., T. Hesson, et al. (1997). "Structure of the hepatitis C virus RNA helicase domain." Nat Struct Biol **4**(6): 463-467.

Yarranton, G. T. and M. L. Gefter (1979). "Enzyme-catalyzed DNA unwinding: studies on Escherichia coli rep protein." Proc Natl Acad Sci U S A **76**(4): 1658-1662.

Ye, J., A. R. Osborne, et al. (2004). "RecA-like motor ATPases--lessons from structures." Biochim Biophys Acta **1659**(1): 1-18.

Yoder-Hill, J., A. Pause, et al. (1993). "The p46 subunit of eukaryotic initiation factor (eIF)-4F exchanges with eIF-4A." J Biol Chem **268**(8): 5566-5573.

Yu, H. (1999). "Extending the size limit of protein nuclear magnetic resonance." Proc Natl Acad Sci U S A **96**(2): 332-334.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Res **31**(13): 3406-3415.

WEBSITES

BLAST: http://blast.ncbi.nlm.nih.gov/Blast.cgi

Clustal: http://www.clustal.org/

Codon Usage Calculator: http://www.protocol-online.org/tools/sms2/codon_usage.html

Diamond: http://www.diamond.ac.uk/Home.html

ExPASy LALIGN: http://www.ch.embnet.org/software/LALIGN_form.html

ExPASy Protpram: http://web.expasy.org/protparam/

NCBI GenBank: http://www.ncbi.nlm.nih.gov/genbank/

OligoAnalyzer: http://eu.idtdna.com/analyzer/applications/oligoanalyzer/default.aspx

Plasmodium Database: http://plasmodb.org/plasmo/

PubMed: http://www.ncbi.nlm.nih.gov/pubmed

APPENDICES

Appendix 1: Oligonucleotides

Number	Name	Sequence
1	X12-1	5′-GACGCTGCCGAATTCTGGCTTGCTAGGACATCTTTGCCCACGTTGACCCG-3′
2	X12-2	5′-CGGGTCAACGTGGGCAAAGATGTCCTAGCAATGTAATCGTCTATGACGTC-3′
3	X12-3	5′-GACGTCATAGACGATTACATTGCTAGGACATGCTGTCTAGAGACTATCGC-3′
4	X12-4	5′-GCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGTC-3′
5	BUB4	5′-CGGGTCAACGTGGGCAAAGATAGAATAGCAAGCCAGAATTCGGCAGCGTC-3′
6	BUB12	5′-CGGGTCAACGTGGGCAAAGTCTATGCTTAGCGCCAGAATTCGGCAGCGTC-3′
7	3’ Flap A	5′-GCTGTCTAGAGACTATCGC-3′
8	5’ Flap A	5′-CGGGTCAACGTGGGCAAAG-3′
9	BF18	5′-GCCTCGCTGCCGTCGCCA-3′
10	BF36	5′-TGGCGACGGCAGCGAGGCTTTTTTTTTTTTTTTTT-3′
11	JS-80	5′-CCAAAAGCACCACACCCCACGCAAAAACAAGTTTTTGCTGATTTTTCTTTATAAATAGAGTGTTATGAAAAATTAGTTTC-3′
12	JS-80nic	5′-GAAACTAATTTTTCATAACACTCTATTTATAAAGAAAAATCAGCAAAAACTTGTTTTTGCGTGGGGTGTGGTGCTTTTGG-3′
13	5’ Hairpin A	5′-GATCCGACCGGCTACTCTAATAGCCGGTTGGACGCACATACTGTGTGCATATG-3′

14	3’ Flap B	5′-CATATGCACACAGTATGTGCGTCCAG-3′
15	5’ Hairpin B	5′-GATCCGATCGGCCATTAGAGTGGCCGATTGGACGCACATACTGTGTGCATATG-3′
16	Deleted Loop	5′-GATCCGATCGGCCAGTGGCCGATTGGACGCACATACTGTGTGCATATG-3′
17	3’ Blunt	5′-CATATGCACACAGTATGTGCGTCCA-3′
18	5’ Hairpin C	5′-ACCGGCTACTCTAATAGCCGGTTGGACGCACATACTGTGTGCATATG-3′
19	3’ Hairpin A	5′-GATCCATATGCACACAGTATGTGCGTCCAACCGGCTATTAGAGTAGCCGGTTTGATCC-3′
20	5’ Flap B	5′-GTGGACGCACATACTGTGTGCATATGGATC-3′
21	5’ Blunt	5′-TGGACGCACATACTGTGTGCATATGGATC-3′
22	3’ Hairpin B	5′-GATCCATATGCACACAGTATGTGCGTCCAACCGGCTATTAGAGTAGCCGGT-3′
23	Rev 3’ Flap A	5′-CGCTATCAGAGATCTGTCG-3′
24	Rev X12-4	5′-CTGCGACGGCTTAAGACCGAACGATCCTGTACGACAGATCTCTGATAGCG-3′
25	26 Overhang	5′-GCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCA-3′
26	21 Overhang	5′-GCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATT-3′
27	16 Overhang	5′-GCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCA-3′
28	11 Overhang	5′-GCGATAGTCTCTAGACAGCATGTCCTAGCA-3′
29	6 Overhang	5′-GCGATAGTCTCTAGACAGCATGTCC-3′
30	1 Overhang	5′-GCGATAGTCTCTAGACAGCA-3′

31	RNA X12-4	5′-GCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGUC-3′
32	RNA X12-1	5′-GACGCUGCCGAAUUCUGGCUUGCUAGGACAUCUUUGCCCACGUUGACCCG-3′
33	Duplex A	5′-AAGAAAAGAAAGAAGAAAGAAACCGTGCAGAATTCGAGGTCGACGGATCCGG-3′
34	Duplex B	5′-CCGGATCCGTCGACCTCGAATTCTGCACGGTTTCTTTCTTCTTTCTTTTCTT-3′
35	Duplex C	5′-CCGTGCAGAATTCGAGGTCGACGGATCCGGAAGAAAAGAAAGAAGAAAGAAA-3′
36	Duplex D	5′-TTTCTTTCTTCTTTCTTTTCTTCCGGATCCGTCGACCTCGAATTCTGCACGG-3′
37	TFO	5′-TTCTTTTCTTTCTTCTTTCTTT-3′
38	X12-4 Comp.	5′-GACGCTGCCGAATTCTGGCTTGCTAGGACATGCTGTCTAGAGACTATCGC-3′

Appendix 2: *PFH45* Sequences

Unspliced Native DNA (PlasmoDB gene ID: PF14_0655)

ATGAGTACTA AAGAAGAGAC TTTTAATAAT GAAAACGACA TTGAAGGAAA TACAGAAGAG
ATTGTTGATA CATTCGACGC TCTTGGATTA AATGAAAAAC TTTTAAGGGG TATATATTCT
TATGGTTTTG AAAAGCCATC AGCAATTCAA CAAAGAGGTA TTAAACCTAT TTTGAATGGT
TACGATACTA TTGGTCAAGC TCAATCTGGT ACAGGAAAAA CAGCTACTTT CGTTATTTCC
TCATTGCAGT TAATTAATTA CGATTATGTT GCATGTCAAG CTTTAATTTT AGCCCCCACT
CGTGAGTTAG CTCAACAAAT TCAAAAGGTA ATAATGAATA AAATGAAATA TAGAAGGTTA
TATATGTAAA TAGATAAATA TATATATATA TATATATATA TATATGTATA TACATATTTA
TGTATGATTT ATGATTTGTT AATTTTTAGG TTGTTTTGGC TTTGGGAGAT TATTTAAAAG
TCAAGTGCCA TGCTTGTGTC GGAGGTACCG TTGTAAGAGA AGATATTGAT AAACTTAAAC
AAGGAGTACA TATGGTTGTT GGTACCCCAG GTAGAGTTTA TGATATGATT GATAAGAGAC
ATTTAGGTGT TGATAGATTA AAGTTATTTA TATTAGATGA AGCAGATGAA ATGTTATCAA
GAGGATTTAA AGCTCAAATA TATGAAGTTT TTAAGAAATT AGTACCAGAT ATTCAAGTTG
CTTTATTTTC TGCTACAATG CCACAAGAAA TATTAGAATT AACTACTAGG TTTATGAGAG
ACCCCAAAAC TATTTTAGTT AAAAAAGATG AATTGACACT TGAAGGTATT AGGCAGTTCT
ATGTTGCTGT AGAAAAGGAA GAATGGAAAT TAGATACCTT ATGTGATTTA TATGAAACAT
TAACCATTAC ACAATCTATA ATATATTGTA ACACTAGGAA AAAGGTTGAT ATATTAACTC
AAGAAATGCA CAATCGTCTT TTCACTGTAT CATGTATGCA CGGAGATATG GACCAAAAAG
ACAGAGATTT AATTATGAGA GAATTCAGAT CTGGATCAAC AAGAGTTTTA GTAACAACAG
ATTTGTTAGC AAGAGGTATT GATGTACAAC AAGTTTCATT AGTTATCAAT TATGATTTAC
CAGCTTCACC AGATACATAT ATTCACAGGT AATATATGGA AAATGAATAT ATATAAATAT
AAATATATAT ATATATATAT ATAAATATAT GTGTATTTTA TTTTTATTTT TTTCTCTTTA
GAATTGGTCG TTCTGGAAGA TTTGGTCGTA AGGGAGTTGC TATCAACTTT GTTACCAACG
ATGATAAGGA AAAAGATAAG TTAAAGAAAA TCGAATCCTA CTATAGTACT CAAATAGAAG
AAATGCCCTT AGAAGTAATA AAAATAAAAT ATATATATAT ATATGTGTAT CACTTTATAT
GTATCTATTG TTTTATAATA ATATATATTT CATTTTAATA TATATATATG TATATATTTT
TTTTTTTTTT TTTTTTTTTT TTCTTTTGTA GGTTGCTGAC TATTTATAA

Spliced Native DNA (PlasmoDB gene ID: PF14_0655)

ATGAGTACTA AAGAAGAGAC TTTTAATAAT GAAAACGACA TTGAAGGAAA TACAGAAGAG
ATTGTTGATA CATTCGACGC TCTTGGATTA AATGAAAAAC TTTTAAGGGG TATATATTCT
TATGGTTTTG AAAAGCCATC AGCAATTCAA CAAAGAGGTA TTAAACCTAT TTTGAATGGT
TACGATACTA TTGGTCAAGC TCAATCTGGT ACAGGAAAAA CAGCTACTTT CGTTATTTCC
TCATTGCAGT TAATTAATTA CGATTATGTT GCATGTCAAG CTTTAATTTT AGCCCCCACT
CGTGAGTTAG CTCAACAAAT TCAAAAGGTT GTTTTGGCTT TGGGAGATTA TTTAAAAGTC

AAGTGCCATG CTTGTGTCGG AGGTACCGTT GTAAGAGAAG ATATTGATAA ACTTAAACAA
GGAGTACATA TGGTTGTTGG TACCCCAGGT AGAGTTTATG ATATGATTGA TAAGAGACAT
TTAGGTGTTG ATAGATTAAA GTTATTTATA TTAGATGAAG CAGATGAAAT GTTATCAAGA
GGATTTAAAG CTCAAATATA TGAAGTTTTT AAGAAATTAG TACCAGATAT TCAAGTTGCT
TTATTTTCTG CTACAATGCC ACAAGAAATA TTAGAATTAA CTACTAGGTT TATGAGAGAC
CCCAAAACTA TTTTAGTTAA AAAAGATGAA TTGACACTTG AAGGTATTAG GCAGTTCTAT
GTTGCTGTAG AAAAGGAAGA ATGGAAATTA GATACCTTAT GTGATTTATA TGAAACATTA
ACCATTACAC AATCTATAAT ATATTGTAAC ACTAGGAAAA AGGTTGATAT ATTAACTCAA
GAAATGCACA ATCGTCTTTT CACTGTATCA TGTATGCACG GAGATATGGA CCAAAAAGAC
AGAGATTTAA TTATGAGAGA ATTCAGATCT GGATCAACAA GAGTTTTAGT AACAACAGAT
TTGTTAGCAA GAGGTATTGA TGTACAACAA GTTTCATTAG TTATCAATTA TGATTTACCA
GCTTCACCAG ATACATATAT TCACAGAATT GGTCGTTCTG GAAGATTTGG TCGTAAGGGA
GTTGCTATCA ACTTTGTTAC CAACGATGAT AAGGAAAAAG ATAAGTTAAA GAAAATCGAA
TCCTACTATA GTACTCAAAT AGAAGAAATG CCCTTAGAAG TTGCTGACTA TTTATAA

Spliced Codon-Optimised DNA

ATGAGCACCA AAGAAGAAAC CTTTAATAAC GAAAATGATA TTGAAGGCAA ACCGAAGAA
ATTGTGGATA CCTTTGATGC ACTGGGCCTG AATGAAAAAC TGCTGCGTGG TATTTATAGC
TATGGCTTTG AAAAACCGAG CGCAATTCAG CAGCGTGGTA TTAAACCGAT TCTGAATGGC
TATGATACCA TTGGTCAGGC ACAGAGCGGC ACCGGTAAAA CCGCAACCTT TGTTATTAGC
AGCCTGCAGC TGATTAATTA TGATTATGTT GCATGCCAGG CCCTGATTCT GGCACCGACC
CGTGAACTGG CACAGCAGAT TCAGAAAGTT GTTCTGGCAC TGGGTGATTA TCTGAAAGTT
AAATGTCATG CATGTGTTGG TGGCACCGTT GTTCGTGAAG ATATTGATAA ACTGAAACAG
GGTGTGCATA TGGTTGTTGG TACACCGGGT CGTGTTTATG ATATGATTGA TAAACGTCAT
CTGGGTGTGG ATCGTCTGAA ACTGTTTATT CTGGATGAAG CAGATGAAAT GCTGAGCCGT
GGTTTTAAAG CCCAGATTTA TGAAGTGTTT AAAAAACTGG TGCCGGATAT TCAGGTTGCA
CTGTTTAGCG CAACCATGCC GCAAGAAATT CTGGAACTGA CCACCCGTTT TATGCGTGAT
CCGAAAACCA TTCTGGTGAA AAAAGATGAA CTGACCCTGG AAGGTATTCG TCAGTTTTAT
GTTGCCGTGG AAAAAGAAGA ATGGAAACTG GATACCCTGT GCGATCTGTA TGAAACCCTG
ACCATTACCC AGAGCATTAT TTATTGCAAT ACCCGCAAAA AAGTGGATAT TCTGACCCAA
GAAATGCATA ATCGTCTGTT TACCGTTAGC TGTATGCATG GTGATATGGA TCAGAAAGAT
CGCGATCTGA TTATGCGTGA ATTTCGTAGC GGTAGCACCC GTGTTCTGGT TACCACCGAT
CTGCTGGCAC GTGGTATTGA TGTTCAGCAG GTTAGCCTGG TTATTAATTA TGATCTGCCG
GCAAGTCCGG ATACCTATAT TCATCGTATT GGTCGTAGCG GTCGTTTTGG TCGTAAAGGT
GTTGCCATTA ATTTTGTGAC CAATGATGAT AAAGAAAAAG ACAAACTGAA AAAAATCGAA
AGCTATTATA GCACCCAGAT TGAAGAAATG CCGCTGGAAG TTGCAGATTA TCTGTAA

Appendix 3: *PFH45* Codon Optimisation Table

Amino Acid	3 Letter Code	*E. coli* Codon Usage for PFH45 Expression			*P. falciparum* Codon Usage for PFH45 Expression		
		Number	Frequency (‰)	Fraction	Number	Frequency (‰)	Fraction
Ala	GCG	0	0	0	0	0	0
Ala	GCA	15	37.59	0.79	4	10.03	0.21
Ala	GCT	0	0	0	14	35.09	0.74
Ala	GCC	4	10.03	0.21	1	2.51	0.05
Cys	TGT	3	7.52	0.5	5	12.53	0.83
Cys	TGC	3	7.52	0.5	1	2.51	0.17
Asp	GAT	30	75.19	0.97	25	62.66	0.81
Asp	GAC	1	2.51	0.03	6	15.04	0.19
Glu	GAG	0	0	0	3	7.52	0.11
Glu	GAA	28	70.18	1	25	62.66	0.89
Phe	TTT	14	35.09	1	9	22.56	0.64
Phe	TTC	0	0	0	5	12.53	0.36
Gly	GGG	0	0	0	0	0	0
Gly	GGA	0	0	0	11	27.57	0.44
Gly	GGT	19	47.62	0.76	14	35.09	0.56
Gly	GGC	6	15.04	0.24	0	0	0
His	CAT	6	15.04	1	3	7.52	0.5
His	CAC	0	0	0	3	7.52	0.5
Ile	ATA	0	0	0	8	20.05	0.26
Ile	ATT	30	75.19	0.97	20	50.13	0.65
Ile	ATC	1	2.51	0.03	3	7.52	0.1
Lys	AAG	0	0	0	12	30.08	0.41
Lys	AAA	29	72.68	1	17	42.61	0.59
Leu	TTG	0	0	0	6	15.04	0.15
Leu	TTA	0	0	0	30	75.19	0.73
Leu	CTG	41	102.76	1	0	0	0
Leu	CTA	0	0	0	0	0	0
Leu	CTT	0	0	0	5	12.53	0.12
Leu	CTC	0	0	0	0	0	0
Met	ATG	11	27.57	1	11	27.57	1
Asn	AAT	11	27.57	0.92	8	20.05	0.67
Asn	AAC	1	2.51	0.08	4	10.03	0.33
Pro	CCG	10	25.06	1	0	0	0
Pro	CCA	0	0	0	6	15.04	0.6
Pro	CCT	0	0	0	1	2.51	0.1
Pro	CCC	0	0	0	3	7.52	0.3
Gln	CAG	18	45.11	0.9	2	5.01	0.1
Gln	CAA	2	5.01	0.1	18	45.11	0.9
Arg	AGG	0	0	0	4	10.03	0.18
Arg	AGA	0	0	0	14	35.09	0.64

Arg	CGG	0	0	0	0	0	0
Arg	CGA	0	0	0	0	0	0
Arg	CGT	20	50.13	0.91	4	10.03	0.18
Arg	CGC	2	5.01	0.09	0	0	0
Ser	AGT	1	2.51	0.06	2	5.01	0.12
Ser	AGC	16	40.1	0.94	0	0	0
Ser	TCG	0	0	0	0	0	0
Ser	TCA	0	0	0	7	17.54	0.41
Ser	TCT	0	0	0	6	15.04	0.35
Ser	TCC	0	0	0	2	5.01	0.12
Thr	ACG	0	0	0	0	0	0
Thr	ACA	1	2.51	0.03	12	30.08	0.41
Thr	ACT	0	0	0	12	30.08	0.41
Thr	ACC	28	70.18	0.97	5	12.53	0.17
Val	GTG	9	22.56	0.3	0	0	0
Val	GTA	0	0	0	7	17.54	0.23
Val	GTT	21	52.63	0.7	21	52.63	0.7
Val	GTC	0	0	0	2	5.01	0.07
Trp	TGG	1	2.51	1	1	2.51	1
Tyr	TAT	16	40.1	1	13	32.58	0.81
Tyr	TAC	0	0	0	3	7.52	0.19
End	TGA	0	0	0	0	0	0
End	TAG	0	0	0	0	0	0
End	TAA	1	2.51	1	1	2.51	1

Appendix 4: *PFDH60* Sequences

Native DNA (PlasmoDB gene ID: PFL1310c)

ATGAGAAGAA ATTTACTGAA GTTAACCTCT ATAAATTGTA AGAAGAATGA CTATGTTTCC
CTTTCTAACA ATTTTTTTTT AATAAAAAAT AGGTCATATA AATATATAAC CGTAAGAAAC
AAAATAGGAA GAAATGTATT TAATGAGAAG GATAATTGTG CCATATTGAA TCGAGGACAA
ACACATGTCG TAGGTATCAA GAAAAATATA TGTACCGAAA ATAAAGTTAA ACAACAGGAT
GGAGAGTTTT ATATGAATCA AAGAGTTCCA GATGATATGG ATAATGATAT GGATAATAAT
GTGGATAATG ATATGGATAA CAATGTGGAT AATCATGTGG ATAATCATGT GGATAATCAT
GTGGATAATC ATGTGGATAA TCATATGGAT AATCATATAA ATAATCATAC ACATAAAAAT
GTGCATGATG ACACATCATA TGATAATAAT TCTTTCGATA GATACCAAAG AAAAAGAGAC
GAAATAGGTG AAAACGGTTT TGTTCAAAAT AGAAGATATG AAAATTATGA AAAACGTAAA
CAATCAAAAA CAAATTATTC ATATGATAAT GATTATAATC AGGGTTCAAG AAATAAAAGA
AATATGTCAT ATCCTGATAG ATATGAAGAA TTAGGAAGTC ATTTAAAGGA CATCGAATGG
GATAATGTGA AATATAAAAT GGAAAGGCAA AATCTATACA ACAACAATTA TAATAAGTGT
AATGATCAGA ATAATAATCA GAATAGTAAT GATAATTTAA ATAATGAGCA AACAAATTGT
TTAAGTAAAG AGGATATACA AAACGAATTA AAGAAAAATA ATATTTATAT TAACAAGGAT
GGAATAATAC ATAATATCAT TAATAAATTC TCAGATGTAT GTTTTCATGA ATCCATATTA
AATTATCTAA ATAATAAATT TAGTGAACCT ACAGCTATTC AAAAAATCAC ATGGCCTATT
GCTTTATCAG GAAAAGATCT TATAGGTGTA GCAGAAACGG GAAGTGGAAA AACGTTAGCT
TTTGTTTTGC CGTGTTTTAT GCATATACTT AAGCATAAAG AAATTATACA AAATAATAAT
AATAATAAAC AAACAAATCA AAACAATAAT TTACAGCAAG AACAGGAAGG GAAAACAAAT
AATCATCATA TAATAAATAA TTATAATGAA AATAATGACA ACAATGAAAA TAATCACAAC
AACAACAACA ATAATAATAA TAATAATAAT AATAGTTATG ATTATGATTA TGATATAAAT
TCCTCTGAAC GGGCAAGTGA TACTTATGGA CTCATTTTAT TACCCACGAG AGAATTGTGC
TTACAAGTTT TAGATGAAAT CAAATCCTTC GAAAAAAACC TACCAATTAA AAGTGTTGCT
GTTTATGGTG GAGTACCGAA ATATTATCAA ATAAATAATT TAAAAAAAGG TGCTGATATT
ATTGTTGCTA CCCCAGGAAG ACTTTTAGAC TTTTTAGAAA ATGGTAACAT CAACCTTTTA
AAATGTATTT ATGTTGTTAT TGATGAAGCT GATCGTTTAC TTGATATGGG ATTCGAAAAA
CAATTAAGAA AAATTATGAC CCAAGTTAAT AAAAATAAAC AATTGCTATT CTTAACTGCT
ACATGGCCTG AACAAGTCAG GAAACTTGCA TATGATTTTT GTTCCTATGA TCCAGTAAAA
ATTCAAATAG GAAAAAATGA ATTGACAGCA AATAAAAATA TTGAGCAAAA TGTTATTATA
AGTTCTTCTA TTGATATGAA GAAAAAATTA TTAGATTGGT TAAAAGAAAA TTATGAAAAT
AATAAAATAT TAATATTTTG TGATACCAAA AGAAATTGTG ATAATTTATG TAAAGAATTA
AGATATCATC AGTATAATGC TTTATCAATT CATGGGGATA AGCAGCAAAG AGAAAGAGAT
CGAATATTGA ATAATTACAA AACAGATCGA TGTAATATAT TAGTAGCAAC CGATGTAGCT
TCAAGAGGAT TAGATATTAA AAATATTTCT GTTGTCATTA ATTATGATAT ACCTAATACT
ATAGAAGATT ATATACATAG AATAGGAAGA ACAGGAAGAG CTGGGAAAAA AGGGAAATCA
ATCTTGTTCT TTTCGTATGA TTATTATATG CCACAAAAAT TAAAGTTTGC AAAAGAGTTA

ATTAAATTGT TAAATAAAAC AAATCAAACT GTTCCACCAC AGTTAAAAGA AATAGCTTAT TCAAGATGA

Codon-Optimised DNA

ATGCGTCGTA ATCTGCTGAA ACTGACCAGC ATTAATTGCA AAAAAAATGA CTACGTGAGC
CTGAGCAACA ACTTTTTTCT GATTAAAAAT CGCAGCTATA AATATATTAC CGTGCGCAAT
AAAATTGGTC GCAACGTGTT TAACGAAAAA GATAATTGCG CCATTCTGAA TCGTGGTCAG
ACCCATGTTG TGGGCATTAA AAAAAACATC TGCACCGAAA ATAAAGTGAA ACAGCAGGAT
GGCGAGTTCT ATATGAATCA GCGTGTTCCG GATGATATGG ATAATGATAT GGACAACAAC
GTGGATAACG ACATGGATAA TAACGTTGAC AACCATGTGG ATAACCATGT TGATAATCAC
GTGGACAATC ACGTCGATAA TCATATGGAT AACCATATCA ACAATCATAC CCATAAAAAC
GTGCACGATG ATACCAGCTA TGATAACAAT AGCTTCGATC GTTATCAGCG CAAACGTGAT
GAAATTGGTG AAAATGGCTT TGTTCAGAAC CGTCGCTATG AAAATTATGA GAAACGCAAA
CAGAGCAAAA CCAACTATAG CTATGATAAT GACTACAATC AGGGCAGCCG CAATAAACGT
AATATGAGCT ATCCGGATCG CTATGAAGAA CTGGGTAGCC ATCTGAAAGA TATTGAATGG
GACAATGTGA AATATAAAAT GGAACGCCAG AATCTGTATA ACAATAATTA CAACAAATGC
AATGACCAGA ATAATAATCA GAATAGCAAT GATAATCTGA ATAATGAACA GACCAATTGC
CTGAGCAAAG AAGATATTCA GAACGAACTG AAAAAAAACA ACATCTATAT TAACAAAGAT
GGCATCATCC ACAATATCAT TAATAAATTT AGCGACGTGT GCTTTCACGA AAGCATTCTG
AATTATCTGA ATAACAAATT CAGCGAACCG ACCGCCATTC AGAAAATTAC CTGGCCGATT
GCACTGAGCG GTAAAGATCT GATTGGTGTT GCAGAAACCG GTAGCGGTAA AACCCTGGCA
TTTGTTCTGC CGTGTTTTAT GCATATTCTG AAACATAAAG AAATCATTCA GAACAACAAT
AATAACAAAC AGACCAATCA GAACAATAAT CTGCAGCAAG AACAAGAAGG CAAAACCAAC
AACCATCACA TCATCAACAA TTACAATGAA AATAACGACA ACAACGAAAA TAATCATAAC
AACAACAACA ACAATAACAA TAACAACAAT AATAGCTATG ACTATGACTA TGATATTAAT
AGCAGCGAAC GTGCAAGCGA TACCTATGGT CTGATTCTGC TGCCGACCCG TGAACTGTGT
CTGCAGGTTC TGGATGAAAT CAAAAGCTTT GAAAAAAATC TGCCGATCAA AAGCGTTGCA
GTTTATGGTG GTGTGCCGAA ATATTATCAG ATTAATAATC TGAAAAAAGG TGCCGATATT
ATTGTTGCAA CACCGGGTCG TCTGCTGGAT TTTCTGGAAA ATGGCAATAT CAACCTGCTG
AAATGCATCT ATGTGGTTAT TGATGAAGCA GATCGCCTGC TGGATATGGG TTTTGAAAAA
CAGCTGCGCA AAATTATGAC CCAGGTGAAT AAAAACAAAC AGCTGCTGTT TCTGACCGCA
ACCTGGCCTG AACAGGTTCG TAAACTGGCA TATGATTTCT GTAGCTATGA CCCGGTGAAA
ATTCAGATTG GCAAAAATGA ACTGACCGCC AACAAAAACA TCGAACAGAA TGTGATTATT
AGCAGCAGCA TTGATATGAA AAAAAAACTG CTGGATTGGC TGAAAGAAAA TTACGAGAAT
AATAAAATTC TGATTTTTTG CGATACCAAA CGCAACTGTG ACAATCTGTG TAAAGAACTG
CGCTATCATC AGTATAATGC CCTGAGCATT CATGGTGATA AACAGCAGCG TGAACGTGAT
CGTATTCTGA ATAACTATAA AACCGACCGC TGCAATATTC TGGTTGCAAC CGATGTTGCA
AGCCGTGGTC TGGATATCAA AAATATTAGC GTGGTGATTA ATTATGATAT TCCGAACACC

ATTGAAGATT ACATCCATCG TATTGGTCGT ACCGGTCGTG CAGGTAAAAA AGGTAAAAGC
ATCCTGTTCT TTAGCTATGA TTACTACATG CCGCAGAAAC TGAAATTTGC CAAAGAACTG
ATTAAACTGC TGAATAAAAC CAATCAGACC GTTCCGCCTC AGCTGAAAGA AATTGCATAT
AGCCGCTAA

Appendix 5: *PFDH60* Codon Optimisation Table

Amino Acid	3 Letter Code	*E. coli* Codon Usage for PFDH60 Expression			*P. falciparum* Codon Usage for PFDH60 Expression		
		Number	Frequency (‰)	Fraction	Number	Frequency (‰)	Fraction
Ala	GCG	0	0	0	0	0	0
Ala	GCA	13	17.5	0.68	6	8.08	0.32
Ala	GCT	0	0	0	12	16.15	0.63
Ala	GCC	6	8.08	0.32	1	1.35	0.05
Cys	TGT	5	6.73	0.36	13	17.5	0.93
Cys	TGC	9	12.11	0.64	1	1.35	0.07
Asp	GAT	43	57.87	0.74	52	69.99	0.9
Asp	GAC	15	20.19	0.26	6	8.08	0.1
Glu	GAG	3	4.04	0.08	6	8.08	0.15
Glu	GAA	36	48.45	0.92	33	44.41	0.85
Phe	TTT	15	20.19	0.75	14	18.84	0.7
Phe	TTC	5	6.73	0.25	6	8.08	0.3
Gly	GGG	0	0	0	4	5.38	0.14
Gly	GGA	0	0	0	16	21.53	0.57
Gly	GGT	20	26.92	0.71	8	10.77	0.29
Gly	GGC	8	10.77	0.29	0	0	0
His	CAT	15	20.19	0.71	20	26.92	0.95
His	CAC	6	8.08	0.29	1	1.35	0.05
Ile	ATA	0	0	0	30	40.38	0.51
Ile	ATT	42	56.53	0.71	22	29.61	0.37
Ile	ATC	17	22.88	0.29	7	9.42	0.12
Lys	AAG	0	0	0	13	17.5	0.19
Lys	AAA	68	91.52	1	55	74.02	0.81
Leu	TTG	0	0	0	8	10.77	0.14
Leu	TTA	0	0	0	36	48.45	0.63
Leu	CTG	57	76.72	1	1	1.35	0.02
Leu	CTA	0	0	0	4	5.38	0.07
Leu	CTT	0	0	0	7	9.42	0.12
Leu	CTC	0	0	0	1	1.35	0.02
Met	ATG	13	17.5	1	13	17.5	1
Asn	AAT	74	99.6	0.6	106	142.66	0.86
Asn	AAC	49	65.95	0.4	17	22.88	0.14
Pro	CCG	13	17.5	0.87	2	2.69	0.13
Pro	CCA	0	0	0	7	9.42	0.47
Pro	CCT	2	2.69	0.13	5	6.73	0.33
Pro	CCC	0	0	0	1	1.35	0.07
Gln	CAG	32	43.07	0.94	9	12.11	0.26
Gln	CAA	2	2.69	0.06	25	33.65	0.74
Arg	AGG	0	0	0	3	4.04	0.09
Arg	AGA	0	0	0	25	33.65	0.74

Arg	CGG	0	0	0	1	1.35	0.03
Arg	CGA	0	0	0	3	4.04	0.09
Arg	CGT	19	25.57	0.56	2	2.69	0.06
Arg	CGC	15	20.19	0.44	0	0	0
Ser	AGT	0	0	0	9	12.11	0.26
Ser	AGC	34	45.76	1	0	0	0
Ser	TCG	0	0	0	1	1.35	0.03
Ser	TCA	0	0	0	12	16.15	0.35
Ser	TCT	0	0	0	7	9.42	0.21
Ser	TCC	0	0	0	5	6.73	0.15
Thr	ACG	0	0	0	3	4.04	0.11
Thr	ACA	1	1.35	0.04	14	18.84	0.5
Thr	ACT	0	0	0	4	5.38	0.14
Thr	ACC	27	36.34	0.96	7	9.42	0.25
Val	GTG	18	24.23	0.51	8	10.77	0.23
Val	GTA	0	0	0	9	12.11	0.26
Val	GTT	16	21.53	0.46	15	20.19	0.43
Val	GTC	1	1.35	0.03	3	4.04	0.09
Trp	TGG	4	5.38	1	4	5.38	1
Tyr	TAT	31	41.72	0.79	36	48.45	0.92
Tyr	TAC	8	10.77	0.21	3	4.04	0.08
End	TGA	0	0	0	1	1.35	1
End	TAG	0	0	0	0	0	0
End	TAA	1	1.35	1	0	0	0

Appendix 6: BLAST Alignment of PFH45 against EIF4A1

ALIGN calculates a global alignment of two sequences version 2.2u
Please cite: Myers and Miller, CABIOS (1989) 4:11-17
PFH45 398 bp 398 aa vs.
EIF4A1 406 bp 406 aa
using matrix file: BLOSUM50, gap open/ext: -14/-4
66.7% identity in 408 aa overlap; Global score: 1786

```
                        10        20        30        40        50
PFH45  ---------MSTKEETFNNENDIEGNTEEIVDTFDALGLNEKLLRGIYSYGFEKPSAIQQ
                 ..  . .. :. ::.: .:::::.:: ..:.:.:::::::::::::::::
EIF4A1 MSASQDSRSRDNGPDGMEPEGVIESNWNEIVDSFDDMNLSESLLRGIYAYGFEKPSAIQQ
                10        20        30        40        50        60

               60        70        80        90       100       110
PFH45  RGIKPILNGYDTIGQAQSGTGKTATFVISSLQLINYDYVACQALILAPTRELAQQIQKVV
       :.: : ..:::.:.:::::::::::::.:: :: :. :  : :::.:::::::::::::::
EIF4A1 RAILPCIKGYDVIAQAQSGTGKTATFAISILQQIELDLKATQALVLAPTRELAQQIQKVV
                70        80        90       100       110       120

             120       130       140       150       160       170
PFH45  LALGDYLKVKCHACVGGTVVREDIDKLK-QGVHMVVGTPGRVYDMIDKRHLGVDRLKLFI
       .:::::. ..::::.::: :: ...::. .. :..::::::::.::...:.:.   .:.:.
EIF4A1 MALGDYMGASCHACIGGTNVRAEVQKLQMEAPHIIVGTPGRVFDMLNRRYLSPKYIKMFV
               130       140       150       160       170       180

               180       190       200       210       220       230
PFH45  LDEADEMLSRGFKAQIYEVFKKLVPDIQVALFSATMPQEILELTTRFMRDPKTILVKKDE
       :::::::::::: :::..:.::   . ::.:.::::::..::.: .:::::  ::::::.:
EIF4A1 LDEADEMLSRGFKDQIYDIFQKLNSNTQVVLLSATMPSDVLEVTKKFMRDPIRILVKKEE
               190       200       210       220       230       240

               240       250       260       270       280       290
PFH45  LTLEGIRQFYVAVEKEEWKLDTLCDLYETLTITQSIIYCNTRKKVDILTQEMHNRLFTVS
       ::::::::::. ::.:::::::::::::::::::::..:. :::.::: ::..:: : ::::
EIF4A1 LTLEGIRQFYINVEREEWKLDTLCDLYETLTITQAVIFINTRRKVDWLTEKMHARDFTVS
               250       260       270       280       290       300

               300       310       320       330       340       350
PFH45  CMHGDMDQKDRDLIMREFRSGSTRVLVTTDLLARGIDVQQVSLVINYDLPASPDTYIHRI
        ::::::::.::.:::::::::.:::.:::::::::::::::::::::.. ..:::::
EIF4A1 AMHGDMDQKERDVIMREFRSGSSRVLITTDLLARGIDVQQVSLVINYDLPTNRENYIHRI
               310       320       330       340       350       360

               360       370       380       390
PFH45  GRSGRFGRKGVAINFVTNDDKEKDKLKKIESYYSTQIEEMPLEVADYL
       ::.:::::::::::.::..::.   :. ::..:.:.::::::.::: .
EIF4A1 GRGGRFGRKGVAINMVTEEDKRT--LRDIETFYNTSIEEMPLNVADLI
               370       380         390       400
```

Appendix 7: FASTA Amino Acid Sequences of PFEIF4A and Human EIF4A Isoforms 1, 2, and 3

PFEIF4A

MSTKEETFNNENDIEGNTEEIVDTFDALGLNEKLLRGIYSYGFEKPSAIQQRGIKPILNGYDTIGQAQSGTGK
TATFVISSLQLINYDYVACQALILAPTRELAQQIQKVVLALGDYLKVKCHACVGGTVVREDIDKLKQGVHMVV
GTPGRVYDMIDKRHLGVDRLKLFILDEADEMLSRGFKAQIYEVFKKLVPDIQVALFSATMPQEILELTTRFMR
DPKTILVKKDELTLEGIRQFYVAVEKEEWKLDTLCDLYETLTITQSIIYCNTRKKVDILTQEMHNRLFTVSCM
HGDMDQKDRDLIMREFRSGSTRVLVTTDLLARGIDVQQVSLVINYDLPASPDTYIHRIGRSGRFGRKGVAINF
VTNDDKEKDKLKKIESYYSTQIEEMPLEVADYL

EIF4A1 (NCBI accession number: AAV38684.1)

MSASQDSRSRDNGPDGMEPEGVIESNWNEIVDSFDDMNLSESLLRGIYAYGFEKPSAIQQRAILPCIKGYDVI
AQAQSGTGKTATFAISILQQIELDLKATQALVLAPTRELAQQIQKVVMALGDYMGASCHACIGGTNVRAEVQK
LQMEAPHIIVGTPGRVFDMLNRRYLSPKYIKMFVLDEADEMLSRGFKDQIYDIFQKLNSNTQVVLLSATMPSD
VLEVTKKFMRDPIRILVKKEELTLEGIRQFYINVEREEWKLDTLCDLYETLTITQAVIFINTRRKVDWLTEKM
HARDFTVSAMHGDMDQKERDVIMREFRSGSSRVLITTDLLARGIDVQQVSLVINYDLPTNRENYIHRIGRGGR
FGRKGVAINMVTEEDKRTLRDIETFYNTSIEEMPLNVADLI

EIF4A2 (NCBI accession number: AAP88862.1)

MSGGSADYNREHGGPEGMDPDGVIESNWNEIVDNFDDMNLKESLLRGIYAYGFEKPSAIQQRAIIPCIKGYDV
IAQAQSGTGKTATFAISILQQLEIEFKETQALVLAPTRELAQQIQKVILALGDYMGATCHACIGGTNVRNEMQ
KLQAEAPHIVVGTPGRVFDMLNRRYLSPKWIKMFVLDEADEMLSRGFKDQIYEIFQKLNTSIQVVLLSATMPT
DVLEVTKKFMRDPIRILVKKEELTLEGIKQFYINVEREEWKLDTLCDLYETLTITQAVIFLNTRRKVDWLTEK
MHARDFTVSALHGDMDQKERDVIMREFRSGSSRVLITTDLLARGIDVQQVSLVINYDLPTNRENYIHRIGRGG
RFGRKGVAINFVTEEDKRILRDIETFYNTTVEEMPMNVADLI

EIF4A3 (NCBI accession number: AAH11151.1)

MATTATMATSGSARKRLLKEEDMTKVEFETSEEVDVTPTFDTMGLREDLLRGIYAYGFEKPSAIQQRAIKQII
KGRDVIAQSQSGTGKTATFSISVLQCLDIQVRETQALILAPTRELAVQIQKGLLALGDYMNVQCHACIGGTNV

GEDIRKLDYGQHVVAGTPGRVFDMIRRRSLRTRAIKMLVLDEADEMLNKGFKEQIYDVYRYLPPATQVVLISA
TLPHEILEMTNKFMTDPIRILVKRDELTLEGIKQFFVAVEREEWKFDTLCDLYDTLTITQAVIFCNTKRKVDW
LTEKMREANFTVSSMHGDMPQKERESIMKEFRSGASRVLISTDVWARGLDVPQVSLIINYDLPNNRELYIHRI
GRSGRYGRKGVAINFVKNDDIRILRDIEQYYSTQIDEMPMNVADLI

Appendix 8: Clustal Alignment of PFEIF4A with Human EIF4A Isoforms 1, 2, and 3

```
EIF4A1   --MSASQDSRSRDNG-PDGMEPEGVIESNWNEIVDS---FDDMNLSESLLRGIYAYGFEK 54
EIF4A2   --MSGGSADYNREHGGPEGMDPDGVIESNWNEIVDN---FDDMNLKESLLRGIYAYGFEK 55
PFEIF4A  --MSTKEETFNNEND----------IEGNTEEIVDT---FDALGLNEKLLRGIYSYGFEK 45
EIF4A3   MATTATMATSGSARKRLLKEEDMTKVEFETSEEVDVTPTFDTMGLREDLLRGIYAYGFEK 60
            :      .  .           :* : .* **    ** :.* *.******:*****

EIF4A1   PSAIQQRAILPCIKGYDVIAQAQSGTGKTATFAISILQQIELDLKATQALVLAPTRELAQ 114
EIF4A2   PSAIQQRAIIPCIKGYDVIAQAQSGTGKTATFAISILQQLEIEFKETQALVLAPTRELAQ 115
PFEIF4A  PSAIQQRGIKPILNGYDTIGQAQSGTGKTATFVISSLQLINYDYVACQALILAPTRELAQ 105
EIF4A3   PSAIQQRAIKQIIKGRDVIAQSQSGTGKTATFSISVLQCLDIQVRETQALILAPTRELAV 120
         *******.*   ::* *.*.*:********** ** ** :: :    ***:********

EIF4A1   QIQKVVMALGDYMGASCHACIGGTNVRAEVQKLQMEAPHIIVGTPGRVFDMLNRRYLSPK 174
EIF4A2   QIQKVILALGDYMGATCHACIGGTNVRNEMQKLQAEAPHIVVGTPGRVFDMLNRRYLSPK 175
PFEIF4A  QIQKVVLALGDYLKVKCHACVGGTVVREDIDKLKQGV-HMVVGTPGRVYDMIDKRHLGVD 164
EIF4A3   QIQKGLLALGDYMNVQCHACIGGTNVGEDIRKLDYGQ-HVVAGTPGRVFDMIRRRSLRTR 179
         **** ::*****: . ****:*** *  :: **.    *::.******:**: :* *

EIF4A1   YIKMFVLDEADEMLSRGFKDQIYDIFQKLNSNTQVVLLSATMPSDVLEVTKKFMRDPIRI 234
EIF4A2   WIKMFVLDEADEMLSRGFKDQIYEIFQKLNTSIQVVLLSATMPTDVLEVTKKFMRDPIRI 235
PFEIF4A  RLKLFILDEADEMLSRGFKAQIYEVFKKLVPDIQVALFSATMPQEILELTTRFMRDPKTI 224
EIF4A3   AIKMLVLDEADEMLNKGFKEQIYDVYRYLPPATQVVLISATLPHEILEMTNKFMTDPIRI 239
          :*:::********.:*** ***:::: * .  **.*:***:* ::**:*.:** **  *

EIF4A1   LVKKEELTLEGIRQFYINVEREEWKLDTLCDLYETLTITQAVIFINTRRKVDWLTEKMHA 294
EIF4A2   LVKKEELTLEGIKQFYINVEREEWKLDTLCDLYETLTITQAVIFLNTRRKVDWLTEKMHA 295
PFEIF4A  LVKKDELTLEGIRQFYVAVEKEEWKLDTLCDLYETLTITQSIIYCNTRKKVDILTQEMHN 284
EIF4A3   LVKRDELTLEGIKQFFVAVEREEWKFDTLCDLYDTLTITQAVIFCNTKRKVDWLTEKMRE 299
         ***::*******:**:: **:****:*******:******::*: **::*** **::*:

EIF4A1   RDFTVSAMHGDMDQKERDVIMREFRSGSSRVLITTDLLARGIDVQQVSLVINYDLPTNRE 354
EIF4A2   RDFTVSALHGDMDQKERDVIMREFRSGSSRVLITTDLLARGIDVQQVSLVINYDLPTNRE 355
PFEIF4A  RLFTVSCMHGDMDQKDRDLIMREFRSGSTRVLVTTDLLARGIDVQQVSLVINYDLPASPD 344
EIF4A3   ANFTVSSMHGDMPQKERESIMKEFRSGASRVLISTDVWARGLDVPQVSLIINYDLPNNRE 359
           ****.:**** **:*: **:*****::***::**: ***:** ****:****** . :

EIF4A1   NYIHRIGRGGRFGRKGVAINMVTEEDKRTLRDIETFYNTSIEEMPLNVADLI 406
EIF4A2   NYIHRIGRGGRFGRKGVAINFVTEEDKRILRDIETFYNTTVEEMPMNVADLI 407
PFEIF4A  TYIHRIGRSGRFGRKGVAINFVTNDDKEKDK---------------------- 375
EIF4A3   LYIHRIGRSGRYGRKGVAINFVKNDDIRILRDIEQYYSTQIDEMPMNVADLI 411
          *******.**:********:*.::* .  :
```

Appendix 9: FASTA Amino Acid Sequences of PFDH60

MRRNLLKLTSINCKKNDYVSLSNNFFLIKNRSYKYITVRNKIGRNVFNEKDNCAILNRGQTHVVGIKKNICTE
NKVKQQDGEFYMNQRVPDDMDNDMDNNVDNDMDNNVDNHVDNHVDNHVDNHVDNHMDNHINNHTHKNVHDDTS
YDNNSFDRYQRKRDEIGENGFVQNRRYENYEKRKQSKTNYSYDNDYNQGSRNKRNMSYPDRYEELGSHLKDIE
WDNVKYKMERQNLYNNNYNKCNDQNNNQNSNDNLNNEQTNCLSKEDIQNELKKNNIYINKDGIIHNIINKFSD
VCFHESILNYLNNKFSEPTAIQKITWPIALSGKDLIGVAETGSGKTLAFVLPCFMHILKHKEIIQNNNNNKQT
NQNNNLQQEQEGKTNNHHIINNYNENNDNNENNHNNNNNNNNNNNNSYDYDYDINSSERASDTYGLILLPTRE
LCLQVLDEIKSFEKNLPIKSVAVYGGVPKYYQINNLKKGADIIVATPGRLLDFLENGNINLLKCIYVVIDEAD
RLLDMGFEKQLRKIMTQVNKNKQLLFLTATWPEQVRKLAYDFCSYDPVKIQIGKNELTANKNIEQNVIISSSI
DMKKKLLDWLKENYENNKILIFCDTKRNCDNLCKELRYHQYNALSIHGDKQQRERDRILNNYKTDRCNILVAT
DVASRGLDIKNISVVINYDIPNTIEDYIHRIGRTGRAGKKGKSILFFSYDYYMPQKLKFAKELIKLLNKTNQT
VPPQLKEIAYSR

Appendix 10: FASTA Amino Acid Sequences of PFDH60

ALIGN calculates a global alignment of two sequences version 2.2u
Please cite: Myers and Miller, CABIOS (1989) 4:11-17
PFDH60 742 bp 742 aa vs.
ATP-DEPENDENT RNA HELICASE 707 bp 707 aa
using matrix file: BLOSUM50, gap open/ext: -14/-4
30.8% identity in 775 aa overlap; Global score: 822

```
               10        20        30        40        50
PFDH60 MRRNLLKLTSINCKKNDYVSLSNNFFLIKNRSYKYI-TVRNKIGRNVFNEKD--NCAILN
       :   .   . ..     ..:   ..:. .. . . .  . ... ::. .  .  :   .
ATP-DE MFVFFTPGNVLDLALMLFLSQLLSIFITHSMALNPLHRLSSSLPRNLTHSLNITNSLTIP
               10        20        30        40        50        60

        60            70        80         90       100       110
PFDH60 RGQTHVVGIKK----NICTENKVKQQDGEFYMN-QRVPDDMDNDMDNNVDNDMDNNVDNH
       :. :: . : .    .:    ..   .  :    .:: . . . . :  . :  ..:
ATP-DE RNLTHSLTIPSLTIPTITIPRSITIPSFTFISPVNRVTSRVKTGLGNFGNYDNISTVTAP
               70        80        90       100       110       120

             120         130       140       150       160
PFDH60 VD-NHVDNHVDNH--VDNHMDNHINNHTHKNVHDDTSYDNNSFDRYQRKRDEIGEN----
        . ... :   :.  : :.          ... . . .  :.:::         : .
ATP-DE YQVGYTPNLSPNYTTVPNYSPYTTPYTGYSTGYGSGTGYNTGFDTNTGYSTGYGSGNTGY
              130       140       150       160       170       180

           170           180        190       200         210
PFDH60 --GFVQNRRYEN-----YEKRKQSKTN-YSYDNDYNQGSRNKR--NMSYPDRYEELGSHL
         :. ..  :..     : ::.  .   :.: : ::   ...:  : . :        .
ATP-DE STGYGSGTGYNTGYGTGYVKRQGYRPRPYNYANRYNTFRQTRRFYNTNTPLGSPPPHYYP
              190       200       210       220       230       240

         220          230       240        250       260
PFDH60 KDIEWD---NVKYKMERQNLYNNNYNKCN-DQNNNQNSNDNLNNEQTNCLSK--EDIQNE
        ..  :   . ..       :.   .. . :...  . . .. . . .  :.  :.:.
ATP-DE GQFTLDFGVHSRFGDAVGRGYDIPVSSIDWDKEELVEIKKDFYDLSYEADSRPGEEIERI
              250       260       270       280       290       300

     270       280       290       300        310       320
PFDH60 LKKNNIYINKDGIIHNIINKFSDVCFHESILNYLN-NKFSEPTAIQKITWPIALSGKDLI
       :: .:: :. .  . . .. :... :...: : .. ..:.::: :::. :   :.:.:.:
ATP-DE LKAHNIIIEGEHPLPKPVTTFDEAVFNQQIQNIIKESNFTEPTPIQKVGWTSCLTGRDII
              310       320       330       340       350       360

      330       340       350       360       370       380
PFDH60 GVAETGSGKTLAFVLPCFMHILKHKEIIQNNNNNKQTNQNNNLQQEQEGKTNNHHIINNY
       ::.:::::::.:.:: ..:.: .  .
ATP-DE GVSQTGSGKTLTFLLPGLLHLLAQPPV--------------------------------
              370       380

      390       400       410       420       430       440
PFDH60 NENNDNNENNNHNNNNNNNNNNNNSYDYDYDINSSERASDTYGLILLPTRELCLQVLDEIK
                                          ..    ::: ::::::::. .: .
ATP-DE ----------------------------------GTGGPIMLILSPTRELCLQIAEEAR
                                        390       400       410

      450       460       470       480       490       500
```

```
PFDH60 SFEKNLPIKSVAVYGGVPKYYQINNLKKGADIIVATPGRLLDFLENGNINLLKCIYVVID
        . . : .. : .:::. :. :. .:..::.:.::::::::.:: ::.:.: .  : :.:
ATP-DE PYSRLLNLRLVPIYGGASKFAQVRELQNGAEIMVATPGRLLEFLSNGTIKLNRVSYFVMD
            420       430       440       450       460       470

      510       520       530       540       550       560
PFDH60 EADRLLDMGFEKQLRKIMTQVNKNKQLLFLTATWPEQVRKLAYDFCSYDPVKIQIGKNEL
       ::::.:::::: :.::::. :.  ..: :...:::: ....:: .::. . . ::.:  ::
ATP-DE EADRMLDMGFEPQIRKIVGQIRPDRQTLMFSATWPSEIKRLASEFCKANSIYIQVGDLEL
            480       490       500       510       520       530

      570       580       590       600       610       620
PFDH60 TANKNIEQNVIISSSIDMKKKLLDWLKENYENNKILIFCDTKRNCDNLCKELRYHQYNAL
       ::: ::.::: . .: ... ::.:.:     ..:.::: : :   :.: . :::.....
ATP-DE TANPNIRQNVEFPNSYEVRDKLFDFLGSIPPEKKVLIFSDLKSFADQLTSALRYRRFKSA
            540       550       560       570       580       590

      630       640       650       660       670       680
PFDH60 SIHGDKQQRERDRILNNYKTDRCNILVATDVASRGLDIKNISVVINYDIPNTIEDYIHRI
       :.::.: : .:.:::: ...   :.::::::::.:::::::.:. ::: :.:... ::::::
ATP-DE SLHGNKTQAQRERILNMFRSGDVNVLVATDVAARGLDIKDIDYVINLDVPKSLLDYIHRI
            600       610       620       630       640       650

      690       700       710       720       730       740
PFDH60 GRTGRAGKKGKSILFFSYDYYMPQKLKFAKELIKLLNKTNQTVPPQLKEIAYSR-
       ::::::..::.:.:.:  :   : :.::::.: :::.:.::::: :: .:: .
ATP-DE GRTGRGNSKGESLLYFPIDTLTPAKVKFAQDLSKLLSKVNQTVPSQLTQIANNNL
            660       670       680       690       700
```

Appendix 11: FASTA Amino Acid Sequences of DHX16

MATPAGLERWVQDELHSVLGLSERHVAQFLIGTAQRCTSAEEFVQRLRDTDTLDLSGPARDFALRLWNKVPRK
AVVEKPARAAEREARALLEKNRSYRLLEDSEESSEETVSRAGSSLQKKRKKRKHLRKKREEEEEEEASEKGKK
KTGGSKQQTEKPESEDEWERTERERLQDLEERDAFAERVRQRDKDRTRNVLERSDKKAYEEAQKRLKMAEEDR
KAMVPELRKKSRREYLAKREREKLEDLEAELADEEFLFGDVELSRHERQELKYKRRVRDLAREYRAAGEQEKL
EATNRYHMPKETRGQPARAVDLVEEESGAPGEEQRRWEEARLGAASLKFGARDAASQEPKYQLVLEEEETIEF
VRATQLQGDEEPSAPPTSTQAQQKESIQAVRRSLPVFPFREELLAAIANHQVLIIEGETGSGKTTQIPQYLFE
EGYTNKGMKIACTQPRRVAAMSVAARVAREMGVKLGNEVGYSIRFEDCTSERTVLRYMTDGMLLREFLSEPDL
ASYSVVMVDEAHERTLHTDILFGLIKDVARFRPELKVLVASATMDTARFSTFFDDAPVFRIPGRRFPVDIFYT
KAPEADYLEACVVSVLQIHVTQPPGDILVFLTGQEEIEAACEMLQDRCRRLGSKIRELLVLPIYANLPSDMQA
RIFQPTPPGARKVVVATNIAETSLTIEGIIYVLDPGFCKQKSYNPRTGMESLTVTPCSKASANQRAGRAGRVA
AGKCFRLYTAWAYQHELEETTVPEIQRTSLGNVVLLLKSLGIHDLMHFDFLDPPPYETLLLALEQLYALGALN
HLGELTTSGRKMAELPVDPMLSKMILASEKYSCSEEILTVAAMLSVNNSIFYRPKDKVVHADNARVNFFLPGG
DHLVLLNVYTQWAESGYSSQWCYENFVQFRSMRRARDVREQLEGLLERVEVGLSSCQGDYIRVRKAITAGYFY
HTARLTRSGYRTVKQQQTVFIHPNSSLFEQQPRWLLYHELVLTTKEFMRQVLEIESSWLLEVAPHYYKAKELE
DPHAKKMPKKIGKTREELG